

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2004, 05:12:10 ; Search time 5955 Seconds
(without alignments)
11332.511 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaaccatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_man:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1557	100.0	1557	6	AX505141	AX505141 Sequence
2	1557	100.0	1560	8	AF467711	AF467711 Emericell
3	1557	100.0	1800	6	AX505252	AX505252 Sequence
4	1557	100.0	1833	6	AX505251	AX505251 Sequence
5	236.2	15.2	741	11	PM6812G	AL68183 Penicilli
6	183	11.8	3751	8	AK110216	AK110216 Oryza sat
7	167.8	10.8	4310	8	AK119806	AK119806 Oryza sat
8	61	3.9	7218	6	I66494	I66494 Sequence 14
9	59.6	3.8	241859	2	AC136406	AC136406 Rattus no
10	57.2	3.7	214690	2	AC083889	AC083889 Mus muscu
11	57.2	3.7	219559	10	AC087329	AC087329 Mus muscu
12	56.8	3.6	125020	9	AF429315	AF429315 Homo sapi
13	52.2	3.4	227533	2	AC094895	AC094895 Rattus no
14	52	3.3	178757	10	AC122509	AC122509 Mus muscu
15	51.8	3.3	197714	2	AC128967	AC128967 Rattus no
16	51.8	3.3	217985	2	BX649416	BX649416 Danio rer
17	51.6	3.3	264103	2	AC126651	AC126651 Rattus no
18	51	3.3	5452	6	AX646997	AX646997 Sequence
19	51	3.3	298166	2	AC087563	AC087563 Homo sapi
20	50.8	3.3	236054	2	AC114393	AC114393 Rattus no
21	50	3.2	110000	2	BX649525	Continuation (3 of
22	50	3.2	212125	5	BX324212	BX324212 Zebrafish
23	50	3.2	232885	2	AC106163	AC106163 Rattus no
24	50	3.2	233841	2	BX640576	BX640576 Danio rer
25	49.6	3.2	87120	2	AC012225	AC012225 Homo sapi
26	49.4	3.2	33058	2	AC100211	AC100211 Mus muscu
27	49.2	3.2	694	3	DMHSP70AB	Y14393 Drosophila
28	49.2	3.2	1751	10	MUSLRONA	M34398 Mouse loric
29	49.2	3.2	1804	10	BC026781	BC026781 Mus muscu
30	49.2	3.2	1810	10	BC058223	BC058223 Mus muscu
31	49.2	3.2	1863	3	AY052029	AY052029 Drosophil
32	49.2	3.2	6530	6	AR076241	AR076241 Sequence
33	49.2	3.2	55528	3	AC105764	U09189 Mus musculu
34	49.2	3.2	144152	3	AC007725	AC105764 Caenorhab
35	49.2	3.2	182183	3	AC007725	AC007725 Drosophil
36	49.2	3.2	226782	3	AB003693	AC007889 Drosophil
37	49.2	3.2	255454	2	AC138676	AB003693 Drosophil
38	49.2	3.1	189735	2	BX640537	AC138676 Mus muscu
39	49	3.1	209887	2	AC079424	BX640537 Danio rer
40	48.8	3.1	156506	2	AC068038	AC079424 Mus muscu
41	48.6	3.1	194864	10	AL691489	AC068038 Homo sapi
42	48.6	3.1	223879	9	AC008735	AL691489 Mouse DNA
43	48.6	3.1	149626	2	BX649341	AC008735 Homo sapi
44	48.4	3.1	184211	2	BX072578	BX649341 Danio rer
45	48.4	3.1	184211	2	BX072578	BX072578 Danio rer

ALIGNMENTS

RESULT 1
AX505141
LOCUS AX505141 1557 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 7 from Patent WO02057456.
ACCESSION AX505141
VERSION AX505141.1 GI:23386445
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C.,
Trueheart,J., Holtzman,D. and Sherman,A.
TITLE Regulators of fungal gene expression
JOURNAL Patent: WO 02057456-A 7 25-JUL-2002;

FEATURES	Microbia (US)	Location/Qualifiers
Source	1. .1557	
	/organism="synthetic construct"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:32630"	
	/note="Fungal gene"	
ORIGIN		
Query Match	100.0%;	Score 1557; DB 6; Length 1557;
Best Local Similarity	100.0%;	Pred. No. 0;
Matches 1557;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	ATGGATCTCTAGAAACATCCCTCTCGGCTCGTCTACCAAGTCTGCTCTCAAGATCGGCG 60
Db	1	ATGGATCTCTAGAAACATCCCTCTCGGCTCGTCTACCAAGTCTGCTCTCAAGATCGGCG 60
QY	61	CTCTCTCTCTGCTCCCATCTCGAGCATGCCATGCTCAGTACACGATCGACGCTCAG 120
Db	61	CTCTCTCTCTGCTCCCATCTCGAGCATGCCATGCTCAGTACACGATCGACGCTCAG 120
QY	121	TACCCAGTCTCTCAGCGGACACCCCTGCTCTCTGCAACCCCATCATAGCCAGTGCCTC 180
Db	121	TACCCAGTCTCTCAGCGGACACCCCTGCTCTCTGCAACCCCATCATAGCCAGTGCCTC 180
QY	181	GTCCTCTCACTCGTACATGGGCGACCGCGCTACCGGCTGATCTGAACAGGTACCCGCA 240
Db	181	GTCCTCTCACTCGTACATGGGCGACCGCGCTACCGGCTGATCTGAACAGGTACCCGCA 240
QY	241	TCAAGTACAGATGTTACGCGTCTCTGCTCGCGGATTAATGCCCCACACTACCGTGGGC 300
Db	241	TCAAGTACAGATGTTACGCGTCTCTGCTCGCGGATTAATGCCCCACACTACCGTGGGC 300
QY	301	AGTTGCTCCGACATCTTTCCTTCTCATGCCATCCGAGGCGGACAGGCGCAG 360
Db	301	AGTTGCTCCGACATCTTTCCTTCTCATGCCATCCGAGGCGGACAGGCGCAG 360
QY	361	CAATGCCGCACTATCTCTCTCATAGCGTCTCCCGCCCGTTCACGGCTCAGTCG 420
Db	361	CAATGCCGCACTATCTCTCTCATAGCGTCTCCCGCCCGTTCACGGCTCAGTCG 420
QY	421	TACCGCAGCAATTCGCGCGGCGCTCCCGGAGACGCTCTGTCTGACTTCAACAATGGA 480
Db	421	TACCGCAGCAATTCGCGCGGCGCTCCCGGAGACGCTCTGTCTGACTTCAACAATGGA 480
QY	481	CTTCTCTCAGGAGCATTCAGTTATTTCGAGCGGAAGCGCTCAAGTTTGGACCCCGTTGCT 540
Db	481	CTTCTCTCAGGAGCATTCAGTTATTTCGAGCGGAAGCGCTCAAGTTTGGACCCCGTTGCT 540
QY	541	GCGAATGCTGCTGCGCGGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT 600
Db	541	GCGAATGCTGCTGCGCGGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT 600
QY	601	CAGGCGCAGCGGTATCTTCCGAGTGTTCGGGACGCGAACTCCCGGTACACAAATGGT 660
Db	601	CAGGCGCAGCGGTATCTTCCGAGTGTTCGGGACGCGAACTCCCGGTACACAAATGGT 660
QY	661	GTTAAGCGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCCGATGGAAGTTCCTCT 720
Db	661	GTTAAGCGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCCGATGGAAGTTCCTCT 720
QY	721	TGCCGAACTGTAAACAGACTTATCTTATGCAAGCATCTCAAGCGCCATCTCGTACGC 780
Db	721	TGCCGAACTGTAAACAGACTTATCTTATGCAAGCATCTCAAGCGCCATCTCGTACGC 780
QY	781	CACACTGTTGACCGCGGTACATGTGTCTTCCGAGACACCTTCTCTCGAGTGAT 840
Db	781	CACACTGTTGACCGCGGTACATGTGTCTTCCGAGACACCTTCTCTCGAGTGAT 840
QY	841	ATCTGAAACGTCATTTCCAAAAATGCTCAATCAGCGCTGTGTAACCCACCGGAGCAACG 900
Db	841	ATCTGAAACGTCATTTCCAAAAATGCTCAATCAGCGCTGTGTAACCCACCGGAGCAACG 900

CDS
1. 1560
/gene="rfec"
/note="putative transcription factor; contains zinc finger
and C2H2 domain; regulator of FLO11"
/codon_start=1
/product="Rfec"
/protein_id="AAO14631.1"
/db_xref="GI:27450539"
/translation="MDPRNHPSPPTSLPQGSAPLPSAPISMPMPQYTMQPYFVS
QHTLPLQPHSQSPHPSYMGPPYRDLNRPASSHDVYASSAPIMPTVGSLS
PPTSLSHNPQAOQAQSQSPHPSVLPASSAQSYPOPIAPAPRRDRADENNG
LPGASYSNGKQWDPVAANGAAPPYKXDSPTQVVGSGQRRGILPSVPGRAIVT
LNVNGYKNTIIPAKADGKFPNCKNTYLHAKHLRHGHTGDRPIMCVLCKDTF
SRSDILKRHPKQSIIRNGNPTGATHLSHPNHAHVRSQQQAANPVPKQDEVSTVPP
PNIPTGTYGEGVNGVNGAPADPMSVNGMGRGQPEDAFPPGRPHQ
GAWPQAPKQSPVLVQPADPSGHLNIDRNIHQVQVVDPKRPMVPHGPHGEL
DWTSMFQPAPEGVMSQSPGGQEPHIAHVEIKRYPTTTAGQESGNNGLYASTM
SGDGVQPARQ"

ORIGIN
Query Match 100.0%; Score 1557; DB 8; Length 1560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGATCCTAGAAACCATCCCTTCGGCTCCGCTACAGTCTGCTCAGAGTCGGCG 60
Db 1 ATGATCCTAGAAACCATCCCTTCGGCTCCGCTACCACTGCTCCCTCAAGGATCGGCG 60
Qy 61 CCTCTTCCTCTGCTCCCATCTGAGCATGCGCAATGCTCAGTACAGATGCGCCTCAG 120
Db 61 CCTCTTCCTCTGCTCCCATCTGAGCATGCGCAATGCTCAGTACAGATGCGCCTCAG 120
Qy 121 TACCAGTCTCTCAGCGCGCACACCTCGCTCTCTGCAACCCCATCATAGCAGTCGCC 180
Db 121 TACCAGTCTCTCAGCGCGCACACCTCGCTCTCTGCAACCCCATCATAGCAGTCGCC 180
Qy 181 GTCCTCACTGCTACATGCGGAGCGCGCTGACCGCTGATCTGAACAGGTACCCGCA 240
Db 181 GTCCTCACTGCTACATGCGGAGCGCGCTGACCGCTGATCTGAACAGGTACCCGCA 240
Qy 241 TCAAGTCAGCATGTTTACGGTCTCTCTGCGCGGATATGCCACACTACCGTGGC 300
Db 241 TCAAGTCAGCATGTTTACGGTCTCTCTGCGCGGATATGCCACACTACCGTGGC 300
Qy 301 AGCTTGCTTCGACATCTTCTCTCTCATCCCAATCGCAGGCGAGGCACAGCGCAG 360
Db 301 AGCTTGCTTCGACATCTTCTCTCTCATCCCAATCGCAGGCGAGGCACAGCGCAG 360
Qy 361 CAATCGCGCAGCATCTCTCTCTCTCATAGCTGCTCCGCGCGCTTCAGGGCTCAGTCG 420
Db 361 CAATCGCGCAGCATCTCTCTCTCTCATAGCTGCTCCGCGCGCTTCAGGGCTCAGTCG 420
Qy 421 TACCGGAGCAATTCGCGCGCGCTCCCGGAGCGCTGCTGCTGACTCAACAATGGA 480
Db 421 TACCGGAGCAATTCGCGCGCGCTCCCGGAGCGCTGCTGCTGACTCAACAATGGA 480
Qy 481 CTTCCTTCAGGAGCATTCAGTTATTCGACGGAAGCCTCAAGGTTGGGACCCCGTTGCT 540
Db 481 CTTCCTTCAGGAGCATTCAGTTATTCGACGGAAGCCTCAAGGTTGGGACCCCGTTGCT 540
Qy 541 GCGAATGCTGTCGCCGCTATCCCGGGAAGACTCCCGGGAAGCTCCCGGGAAGCTCCCG 600
Db 541 GCGAATGCTGTCGCCGCTATCCCGGGAAGACTCCCGGGAAGCTCCCGGGAAGCTCCCG 600
Qy 601 CAGGGGACCGGCTATCTTCGAGTCTCCGGAGCGCAACTCCGGTCACAAATGGT 660
Db 601 CAGGGGACCGGCTATCTTCGAGTCTTCGGAGCGCAACTCCGGTCACAAATGGT 660
Qy 661 GTTAAAGGACCGGCAAGACACTACTATCCCGGCAAGGATCGGATGGAAGTTCCT 720
Db 661 GTTAAAGGACCGGCAAGACACTACTATCCCGGCAAGGATCGGATGGAAGTTCCT 720
Qy 721 TGCCGGAAGTGAAGAGCTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTAGCG 780

Db 721 TGCCGGAAGTGAAGAGCTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTAGCG 780
Qy 781 CACACTGGTGACCGCGCTACATGTGTCTTTTCAAGACACCTTCTCTCGCAGTCAAT 840
Db 781 CACACTGGTGACCGCGCTACATGTGTCTTTTCAAGACACCTTCTCTCGCAGTCAAT 840
Qy 841 ATCCTGAAACGTCATTTCCAAAATGCTCAATCAGGCGTGTAAACCCACGCGGACCAACG 900
Db 841 ATCCTGAAACGTCATTTCCAAAATGCTCAATCAGGCGTGTAAACCCACGCGGACCAACG 900
Qy 901 CACTTGTGCGCACCCCAATGCGCATGTGAGAGCTCCCAACAGCAGGCTCGCGGAAATCCT 960
Db 901 CACTTGTGCGCACCCCAATGCGCATGTGAGAGCTCCCAACAGCAGGCTCGCGGAAATCCT 960
Qy 961 GTAAAACTCTCAGAGTGAAGTCACTAGTACCGTCCCGCTCCCAATGCAATGCCCGGCG 1020
Db 961 GTAAAACTCTCAGAGTGAAGTCACTAGTACCGTCCCGCTCCCAATGCAATGCCCGGCG 1020
Qy 1021 ACAGTTCAGCGGAGGAGCGCTCAACGCAATGAGTACCTCGGCGCCCGCGCAGGTAC 1080
Db 1021 ACAGTTCAGCGGAGGAGCGCTCAACGCAATGAGTACCTCGGCGCCCGCGCAGGTAC 1080
Qy 1081 GCGGATCACCAGACTATGGGCTTCCCAATGCTCATCCGTCACGCGATGGCGCGTGGTCAG 1140
Db 1081 GCGGATCACCAGACTATGGGCTTCCCAATGCTCATCCGTCACGCGATGGCGCGTGGTCAG 1140
Qy 1141 CCTGAAGACGCGTTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCACAGCTCCC 1200
Db 1141 CCTGAAGACGCGTTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCACAGCTCCC 1200
Qy 1201 AAGCAGAGCCGCTATCTCGTGCAGCGCGTGTGACCCCTTCTGGCCACAGTGAATATT 1260
Db 1201 AAGCAGAGCCGCTATCTCGTGCAGCGCGTGTGACCCCTTCTGGCCACAGTGAATATT 1260
Qy 1261 GACGGAACATCGACGAGTAAACACCGGTTTTCAGACCCCAAGCGCCCTGTGATG 1320
Db 1261 GACGGAACATCGACGAGTAAACACCGGTTTTCAGACCCCAAGCGCCCTGTGATG 1320
Qy 1321 CCAGGACATCCCGGCGCACCCCGGTGAGCTGAGCTGATGTTTCCAACTCAAGCT 1380
Db 1321 CCAGGACATCCCGGCGCACCCCGGTGAGCTGAGCTGATGTTTCCAACTCAAGCT 1380
Qy 1381 CCGAGGGCTACATGTTCTCCAGTCTATGCTGCTGCTCAAGAGCCATCCAGCTCAT 1440
Db 1381 CCGAGGGCTACATGTTCTCCAGTCTATGCTGCTGCTCAAGAGCCATCCAGCTCAT 1440
Qy 1441 GTCGAGCCGAGCGAAGTATTACCCACCACTACCGCTGCTCAAGAGTGGAAATGAAC 1500
Db 1441 GTCGAGCCGAGCGAAGTATTACCCACCACTACCGCTGCTCAAGAGTGGAAATGAAC 1500
Qy 1501 GGTCTCTATCTGGCTTCGATGATGAGTGGCGGACCGGTTCAAGCGCCGAGCAA 1557
Db 1501 GGTCTCTATCTGGCTTCGATGATGAGTGGCGGACCGGTTCAAGCGCCGAGCAA 1557

RESULT 3
AX505252 1800 bp DNA linear PAT 27-SEP-2002
LOCUS Sequence 118 from Patent WO02057456.
DEFINITION AX505252
ACCESSION AX505252
VERSION AX505252.1 GI:23386501
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
ARTIFICIAL SEQUENCES.
REFERENCE 1
AUTHORS Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C.,
Trueheart,J., Holtzman,D. and Sherman,A.
TITLE Regulators of fungal gene expression
JOURNAL Patent: WO 02057456-A 118 25-JUL-2002;
Microbia (US)
FEATURES Location/Qualifiers

source	1. .1800	
	/organism="synthetic construct"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:32630"	
	/note="fungal gene"	
ORIGIN		
Query Match	100.0%; Score 1557; DB 6; Length 1800;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1557; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGGATCTCTAGAAACCATCCCTCTGGCTCCGCTACCAAGTCTGCTCAAGATCGGGC	60
DB	244 ATGGATCTCTAGAAACCATCCCTCTGGCTCCGCTACCAAGTCTGCTCAAGATCGGGC	303
QY	61 CTTCTTCTCTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGAGCCTCAG	120
DB	304 CTTCTTCTCTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGAGCCTCAG	363
QY	121 TACCCAGTCTCTAGCCGACACCCCTGCTCTCTGCAACCCCAATCATAGCAGTGGCC	180
DB	364 TACCCAGTCTCTAGCCGACACCCCTGCTCTCTGCAACCCCAATCATAGCAGTGGCC	423
QY	181 GCTCTCTCACTGATAGTGGGAGCGCGGTACCGGCTTGATCTGACAGGTACCCCGCA	240
DB	424 GCTCTCTCACTGATAGTGGGAGCGCGGTACCGGCTTGATCTGACAGGTACCCCGCA	483
QY	241 TCAAGTCAAGATGTTTACGCGTCTTCTGCTGCGCCGATTAATGCCACACTACCGTGGC	300
DB	484 TCAAGTCAAGATGTTTACGCGTCTTCTGCTGCGCCGATTAATGCCACACTACCGTGGC	543
QY	301 AGCTTGCCTCCGACATCTTTTCCCTTCTCATCCCAATCCGAGCGCAGGACAGGCGCAG	360
DB	544 AGCTTGCCTCCGACATCTTTTCCCTTCTCATCCCAATCCGAGCGCAGGACAGGCGCAG	603
QY	361 CAATCCGCGCACTATCTCTCTCTATAGGTGCTCCGCGCGCTTCCAGCGCTCAGTCG	420
DB	604 CAATCCGCGCACTATCTCTCTCTATAGGTGCTCCGCGCGCTTCCAGCGCTCAGTCG	663
QY	421 TACCCGAGCCAAATGGCGCGCGGCTCTCCCGGACCGTCTGCTGACATTCACCAATGGA	480
DB	664 TACCCGAGCCAAATGGCGCGCGGCTCTCCCGGACCGTCTGCTGACATTCACCAATGGA	723
QY	481 CTTCTCTCAGGAGCATCTAGTTATTCGAGCGGAAAGCTCAAGTTGGGACCCCGTTGCT	540
DB	724 CTTCTCTCAGGAGCATCTAGTTATTCGAGCGGAAAGCTCAAGTTGGGACCCCGTTGCT	783
QY	541 GCGAATGGTGTGCGCGGTATCCGCGGAGGACTCCCGCGACCCAGGTTGTTGTTCT	600
DB	784 GCGAATGGTGTGCGCGGTATCCGCGGAGGACTCCCGCGACCCAGGTTGTTGTTCT	843
QY	601 CAGGGGCGACGCGGTATCTCTCCGAGTGTTCGCGGACCGGCAACTCCGGTCAAAATGGT	660
DB	844 CAGGGGCGACGCGGTATCTCTCCGAGTGTTCGCGGACCGGCAACTCCGGTCAAAATGGT	903
QY	661 GTTAAAGGACCGGAGAACACTATATCCCGGCGAAGATCCCGATGGAATGCCCT	720
DB	904 GTTAAAGGACCGGAGAACACTATATCCCGGCGAAGATCCCGATGGAATGCCCT	963
QY	721 TGCCCGAAGTGTAAAGACTTATCTTCTCATGCCAAGCATCTCAAGGCGCATCTGTACGC	780
DB	964 TGCCCGAAGTGTAAAGACTTATCTTCTCATGCCAAGCATCTCAAGGCGCATCTGTACGC	1023
QY	781 CACATGGTACCGCGCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGAGTGAT	840
DB	1024 CACATGGTACCGCGCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGAGTGAT	1083
QY	841 ATCTGAAAGTGATTTTCCAAAATGCTCAATCAGCGGTGTTAAACCCACCGAGCAACG	900
DB	1084 ATCTGAAAGTGATTTTCCAAAATGCTCAATCAGCGGTGTTAAACCCACCGAGCAACG	1143
QY	901 CACTTGTGCGACCCCAATGGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCAATCTCT	960
DB		
Db	1144 CACTTGTGCGACCCCAATGGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCAATCTCT	1203
QY	961 GTAAAACTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCGGGC	1020
DB	1204 GTAAAACTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCGGGC	1263
QY	1021 AGGACTTACGGGAGGAGCGGTCAACGGCAATGACTAGCTCCGCGCGCGGACAGGTAC	1080
DB	1264 AGGACTTACGGGAGGAGCGGTCAACGGCAATGACTAGCTCCGCGCGCGGACAGGTAC	1323
QY	1081 GGGATACCAAGACTATGCGTTCCTCAATGTCCATCGTCAACGGATGGCGCTGTGTCAG	1140
DB	1324 GGGATACCAAGACTATGCGTTCCTCAATGTCCATCGTCAACGGATGGCGCTGTGTCAG	1383
QY	1141 CTTGAAGACGCGTTCCTCCGCGCGCGCGCATCAAGGAGCCCTTTGGCCACAAGTCC	1200
DB	1384 CTTGAAGACGCGTTCCTCCGCGCGCGCGCATCAAGGAGCCCTTTGGCCACAAGTCC	1443
QY	1201 AAGCAGAGCCGTATCTGTCGAGCGGGTGTGACCCCTTCTGGCCACAGTTGAATAT	1260
DB	1444 AAGCAGAGCCGTATCTGTCGAGCGGGTGTGACCCCTTCTGGCCACAGTTGAATAT	1503
QY	1261 GACCGAAACATCGACAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG	1320
DB	1504 GACCGAAACATCGACAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG	1563
QY	1321 CCAGGACATCCCGGCCACCCCGGTGAGTTGATCGAGCGTCTATGTTTCAACCTCAAGCT	1380
DB	1564 CCAGGACATCCCGGCCACCCCGGTGAGTTGATCGAGCGTCTATGTTTCAACCTCAAGCT	1623
QY	1381 CCCGAGGGTACATCTTCTCCAGTCTATGCTGCTGCTGCTGCTCAAGAGCCCATCCAGCTCAT	1440
DB	1624 CCCGAGGGTACATCTTCTCCAGTCTATGCTGCTGCTGCTCAAGAGCCCATCCAGCTCAT	1683
QY	1441 GTCGAGACCGAGCGAAAGTATTACCCCAACACTACCGTGGTCAAGAGAGTGAATGAAC	1500
DB	1684 GTCGAGACCGAGCGAAAGTATTACCCCAACACTACCGTGGTCAAGAGAGTGAATGAAC	1743
QY	1501 GGTCTCTATCTGGCTTCGACTATGATGGCGAGCGGACCGGTCAGCCCGCCAGACAA	1557
DB	1744 GGTCTCTATCTGGCTTCGACTATGATGGCGAGCGGACCGGTCAGCCCGCCAGACAA	1800
RESULT 4		
AX505251	1833 bp	DNA
LOCUS	Sequence 117 from Patent WO02057456.	linear
DEFINITION		PAT 27-SEP-2002
ACCESSION	AX505251	
VERSION	AX505251.1	GI:23386500
KEYWORDS		
SOURCE	synthetic construct	
ORGANISM	synthetic construct	
REFERENCE	artificial sequences.	
AUTHORS	1	
TITLE	Calif, B.M., Madden, K.T., Milne, T.G., Zhang, L., Silva, J.C.,	
JOURNAL	Trueheart, J., Holtzman, D. and Sherman, A.	
Microbia (US)	Regulators of fungal gene expression	
Patent: WO 02057456-A 117 25-JUL-2002;		
Location/Qualifiers		
1. .1833		
/organism="synthetic construct"		
/mol_type="unassigned DNA"		
/db_xref="taxon:32630"		
/note="fungal gene"		
ORIGIN		
Query Match	100.0%; Score 1557; DB 6; Length 1833;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1557; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGGATCTCTAGAAACCATCCCTCTGGCTCCGCTACCAAGTCTGCTCAAGATCGGGC	60
DB		

QY	721	TGCCCGAAGCTGTACACAGACTTATCTTCATGCCACAGCATCTCAAGCCGCTCTGCTACGC	780
Db	119	TGCGCTCATTTGTATCAACAGAGTACTCTCCACGCGAGCATCTTAAACGCCATCTTTTGAGA	178
QY	781	C-----	ACA 784
Db	179	CGTAAGTTGTTCCACTCAATTCATACCTTTTTCATCTATTACCAATCAACGACGACA	238
QY	785	CTGTGTACCGCCGCTACATGTGTGTTCTTTTGCAAGACACCTTCTCTCGAGTGATATCC	844
Db	239	CCGCTGATCGCCCTACATGTGTCTCTGCAAGATACATCTCCCGAAGCGATATTT	298
QY	845	TGAAAGCTCATTTTCCAAATGCTCAATCAGGCTGTGTATACCCACCGGAGCAAGCACT	904
Db	299	TGAAACGACACTTCCAAAGTGCTGTTAGCGCGCGGTAAACCCACAGGAGCGAGCCATC	358
QY	905	TGTCGACCCCCCAATGCGCATGTGAAGAGGTGCCAACAGCAGGCTCGCGGCAATCTCTTAA	964
Db	359	TCTGCACTCTCAGGCGCACTGAAACGGTC-----TCAGGCGCGAGCTTANGCTGCAA	412
QY	965	AACCTGTCAGAGTGAAGTCACTAGTAGTACCGTCCGCTCCCAATGGCATCCCGGCAACGA	1024
Db	413	AATCCGTTCAAGGAAGTCACTAGTCACTTCCGCTTCCCAATGGTATTGCGGGTACGA	472
QY	1025	CTTACGCGAGGAGCGCTCAACGGCAATGACTAGTCTCGGC-----CGGGCCAG	1075
Db	473	CTTATGGCGAGCGGCGACGCTGAATGGCCATGGCCACACGCTGGCTCGGCGGCGACCTG	532
QY	1076	GGTACGCGGATCACAGCATGTGGCTTCCCAATGTCTCATCGTCAACGGGATGGCGGTG	1135
Db	533	GTCTCAGGAGCAGCAGCTCTGGGATTTCAATGTCTCTGCCACGGCTGAACATG	592
QY	1136	GTCAGCTGTAAGAGCGGTTTCCCGCGCGCGCGCATCAAGAGAGCCCTTGGCCACAAG	1195
Db	593	GTACAGTTGAGAGCGGCTTCAACCTGCCAGGCGAGATCACAGAGCTTCTGGATGCTG	652
QY	1196	CTCCAGAGAGCGGCTATCTGTGAGCGGGGTGCTGACCTTCTGGCC	1246
Db	653	CTCAGAGAGAGCGGCTGCTGTCACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	703
RESULT	6		
LOCUS	AK110216		
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:002-162-D06, full insert sequence.		
ACCESSION	AK110216		
VERSION	AK110216.1 GI:32995425		
KEYWORDS	Full cDNA; oligo capping.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	1		
AUTHORS	The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ohtsuki, S., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Iida, Y., Sugano, S., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariakawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from		

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764
2 (bases 1 to 3751)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K., Imoto, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariakawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ohtsuki, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission

Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kanmoudai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cdna/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

PAIS Genome Sequencing & Analysis Group: Ohtsuki, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariakawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES

source

1. 3751
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="002-162-D06"

ORIGIN

Query Match 11.8%; Score 183; DB 8; Length 3751;
Best Local Similarity 61.0%; Pred. No. 2.3e-35;
Matches 335; Conservative 0; Mismatches 205; Indels 9; Gaps 2;

WEINSTOCK, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 241859)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (01-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241859)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24462281.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: KCQC
Center clone name: CH230-522A18
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 187090 bases at least Q40
Consensus quality: 189991 bases at least Q30
Consensus quality: 191799 bases at least Q20
Estimated insert size: 211470; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 64842: contig of 64842 bp in length
64843 64942: gap of unknown length
64943 70527: contig of 5585 bp in length
70528 70627: gap of unknown length
70628 237692: contig of 167065 bp in length
237693 237792: gap of unknown length
237793 238825: contig of 1033 bp in length
238826 238925: gap of unknown length
238926 241859: contig of 2934 bp in length.

Location/Qualifiers
1. 241859
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-522A18"
2456..3351
/note="clone boundary
clone_end:sp6
site:MboI

misc_feature
source

FEATURES
source

end sequence:BZ285495"
complement(52603..53490)
/note="clone_boundary
clone_end:if7
site:MboI
end sequence:BZ285493"
102059..103179
/note="wgs_end_extension
clone_end:T7"
113425..117805
/note="wgs_end_extension
clone_end:T7"
ORIGIN
Query Match 3.8%; Score 59.6; DB 2; Length 241859;
Best Local Similarity 45.9%; Pred.No. 0.00036;
Matches 203; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
Qy 16 CATCCCTCTCGGCTCGCTTACCAAGTCTGCCTCAGGATCGGCGCTCTTCTCTGCT 75
Db 87310 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87251
Qy 76 CCATCTCGAGCAGCCAAAGCTCTAGTACACAGTACAGCTCAGTACCCAGTCTCTCAG 135
Db 87250 CTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 87191
Qy 136 CCGCACACCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 195
Db 87190 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87131
Qy 196 ATGGGGCAGCGCGGTACCGGCTGTATCTGAACAGGTACCGGATCAAGTACAGTGT 255
Db 87130 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87071
Qy 256 TAGCGTCTCTGCTGCGCGGATAAGCCCGACACTACCGTGGGACGTTGCTCTCGACA 315
Db 87070 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 87011
Qy 316 TCTTTCCTTCTCATCCCAATCGGCGGCGCAGCAGGCGAGCAATCGCCGACTAT 375
Db 87010 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86951
Qy 376 CTTCTCTCTCATAGCGTCTCCCGCGGCTTCAGCGCTCAGTGTACCGGACGCAATT 435
Db 86950 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86891
Qy 436 GCGCGGCGGCTCTCCCGGACC 457
Db 86890 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86869
RESULT 10
AC083889/c
LOCUS
DEFINITION Mus musculus chromosome 5 clone RP23-271A20 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 29 unordered pieces.
AC083889
AC083889.1 GI:10645305
VERSION
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 241859)
REFERENCE
AUTHORS Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Buffard, G.G., Dietrich, N.L., Gupta, J., Ho, S.-L., Iddi, J.,
Lee-Lin, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Tongson, E.E.,
Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A., Wetherby, K.D. and
Green, E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished

misc_feature
note="BZ285495"
complement(52603..53490)
/note="clone_boundary
clone_end:if7
site:MboI
end sequence:BZ285493"
102059..103179
/note="wgs_end_extension
clone_end:T7"
113425..117805
/note="wgs_end_extension
clone_end:T7"
ORIGIN
Query Match 3.8%; Score 59.6; DB 2; Length 241859;
Best Local Similarity 45.9%; Pred.No. 0.00036;
Matches 203; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
Qy 16 CATCCCTCTCGGCTCGCTTACCAAGTCTGCCTCAGGATCGGCGCTCTTCTCTGCT 75
Db 87310 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87251
Qy 76 CCATCTCGAGCAGCCAAAGCTCTAGTACACAGTACAGCTCAGTACCCAGTCTCTCAG 135
Db 87250 CTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 87191
Qy 136 CCGCACACCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 195
Db 87190 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87131
Qy 196 ATGGGGCAGCGCGGTACCGGCTGTATCTGAACAGGTACCGGATCAAGTACAGTGT 255
Db 87130 CCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87071
Qy 256 TAGCGTCTCTGCTGCGCGGATAAGCCCGACACTACCGTGGGACGTTGCTCTCGACA 315
Db 87070 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 87011
Qy 316 TCTTTCCTTCTCATCCCAATCGGCGGCGCAGCAGGCGAGCAATCGCCGACTAT 375
Db 87010 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86951
Qy 376 CTTCTCTCTCATAGCGTCTCCCGCGGCTTCAGCGCTCAGTGTACCGGACGCAATT 435
Db 86950 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86891
Qy 436 GCGCGGCGGCTCTCCCGGACC 457
Db 86890 CCTGTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 86869
RESULT 10
AC083889/c
LOCUS
DEFINITION Mus musculus chromosome 5 clone RP23-271A20 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 29 unordered pieces.
AC083889
AC083889.1 GI:10645305
VERSION
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 241859)
REFERENCE
AUTHORS Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Buffard, G.G., Dietrich, N.L., Gupta, J., Ho, S.-L., Iddi, J.,
Lee-Lin, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Tongson, E.E.,
Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A., Wetherby, K.D. and
Green, E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished

[illegible]

```

# bases
700
600
500
400
300
200
100
0
-----
5 10 15 20 25 30 35 40
Phrap Value Range

```

FEATURES

source

Location/Qualifiers

```

1. .219559
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosomes="5"
/clone="RP23-383N15"
277..346
/rpt_family="CT-rich"
complement(376..513)
/rpt_family="ID_B1"
complement(558..712)
/rpt_family="RSINE1"
complement(1156..1446)
/rpt_family="ID_B1"
1889..1981
/rpt_family="B4"
1982..2201
/rpt_family="B3"
2202..2384
/rpt_family="B4"
complement(2716..2856)
/rpt_family="B1F"
complement(2859..2932)
/rpt_family="ID_B1"
3152..3288
/rpt_family="MT2B"
3178..3316
/rpt_family="B1_MM"
3321..3490
/rpt_family="RSINE1"
3534..3645
/rpt_family="PB1D10"
3875..4023
/rpt_family="B1_MM"
4089..4183
/rpt_family="ZP3AR_MM"
4188..4270
/rpt_family="ID6"
4795..4868
/rpt_family="ID_B1"
4809..4869
/rpt_family="BC1_MM"
4870..5039
/rpt_family="(GGAAA)n"
5040..5062
/rpt_family="BC1_MM"
complement(6119..6242)
/rpt_family="PB1D10"
6339..6482
/rpt_family="RSINE1"
6538..6593
/rpt_family="(CATATA)n"
complement(6846..6990)
/rpt_family="ID_B1"
7322..7385
/rpt_family="A-rich"
complement(7587..7786)
/rpt_family="MER2"
7870..8011

```

```

/rpt_family="B4A"
9384..9471
/rpt_family="PB1D10"
9479..9713
/rpt_family="B4A"
9714..9816
/rpt_family="PB1D9"
9818..9840
/rpt_family="(CAAAA)n"
9841..9865
/rpt_family="B4A"
9866..9889
/rpt_family="(CAAAA)n"
9890..9956
/rpt_family="B4A"
10555..10617
/rpt_family="(CCTG)n"
10731..11005
/rpt_family="B4A"
11064..11279
/rpt_family="B4"
11302..11342
/rpt_family="(CATTA)n"
12080..12111
/rpt_family="AT-rich"
12159..12228
/rpt_family="GA-rich"
12310..12499
/rpt_family="B2_Mm1"
12662..12688
/rpt_family="AT-rich"
12754..12839
/rpt_family="PB1D7"
13195..13270
/rpt_family="ID5"

Query Match      3.7%; Score 57.2; DB 10; Length 219559;
Best Local Similarity 45.7%; Pred.No.0.0015;
Matches 200; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 16 CATCCCTCTCGGGCTCCGCTACCGTCTACCGTCTCGGCTCAGGATCGGGCTCTTCTTCTGCT 75
Db 144415 CCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 144474

QY 76 CCCATCTCGAGCATGCCAATGCTCAGTACACGATCAGCCTCAGTACCGCTCAGTACCGCTCAG 135
Db 144475 CCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 144534

QY 136 CCGCACACCTGCTCTGCTCTGCAACCCCATATAGCCAGTGGCCGCTCTCTACTCGTAC 195
Db 144535 CTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 144594

QY 196 ATGGGGCAGCGCGCTACCGGCTGATCTGAACAGGTACCCGCTACCAAGTACAGTGT 255
Db 144595 CTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 144654

QY 256 TAGCGCTCTTCTGCTGCGCCGATTAATGCGCCACACTACCGTGGGAGCTTGCCTCGACA 315
Db 144655 CCTCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 144714

QY 316 TCTTTCTTTCTCATCCCAATCGGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGTAT 375
Db 144715 CTTCTCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 144774

QY 376 CTTCTCTCTCATAGCGTCTCGCGCGCTTCCAGCGCTCAGTCTGCTACCGCGAGCAATT 435
Db 144775 CTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 144834

QY 436 GCGCGCGCGCTCTCGCGG 453
Db 144835 CTTCTCTGCTCTCTCTG 144852

```

```

RESULT 12
AF429315      125020 bp      DNA      linear      PRI 18-JAN-2002
LOCUS
DEFINITION Homo sapiens junctophilin 3 (JP3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Porter,N.T., Ross,C.A. and Margolis,R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA.
FEATURES
source
1..125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
repeat_region complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
gene complement(<36507..>36887)
/gene="JP3"
/mRNA /note="synonym: JP3"
complement(<36507..>36887)
/gene="JP3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JP3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
CDS /codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSSGRFPDDGGYCGWEDGKAHGHTVCTPKGQGVTSWS
HGFEVLGVYTPSNITQGTWQAKRRIGLGSKRWYKGEWTHGFKRGYVRECG
NGAKYETWSNGLQDGYGTETYSDG"
ORIGIN
Query Match 3.6%; Score 56.8; DB 9; Length 125020;
Best Local Similarity 11.4%; Pred. No. 0.0019;
Matches 108; Conservative 403; Mismatches 422; Indels 13; Gaps 3;
Qy 101 AGTACAGATGACGCTCAGTACAGTCTCTCAGCGGCACACCTGCTCTCTCTGCAAC 160
Db 16874 RRKWKSHAGRRMCCYYSKSWMBMSVSYVKVHSHASBSCHWBKCMTWSCMSNMV 16933
Qy 161 CCATCATAGCAGTCGCCCTCTCTACTGATACATGGGCGCGCGCTGACGGCGCTG 220
Db 16934 KSSWWSGWSGWCGRRRASRKKWYSRGRMSKSMRYTGGSKMRRMCTSSCYASMCC 16993
Qy 221 ATCTGAACAGGTACCCGCATCAGTACAGTATTAGCGGTCTTCTGCTCGCGCGATAA 280
Db 16994 MCCWSCCWRSCCWCMSYCCCRCCYCAKCKYMSYMTMSASYSRMSRYSWKRSKMSRM 17053
Qy 281 TGCCCCACACTACCGTGGCGGAGCTTGCCCTCGACATCTTCTCTCTCATCCCAATCCG 340

```



```

ACCESSION      AC122509
VERSION        AC122509.2  GI:30985117
KEYWORDS       HTG.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus

REFERENCE
AUTHORS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL        1 (bases 1 to 178757)
AUTHORS        Du,H., Boyer,E. and Creason,K.
TITLE          The sequence of Mus musculus BAC clone RP24-443G20
JOURNAL        Unpublished (2001)
REFERENCE      2 (bases 1 to 178757)
AUTHORS        Wilson,R.
TITLE          Sequencing of Mus musculus
JOURNAL        Unpublished (2001)
REFERENCE      3 (bases 1 to 178757)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          Direct Submission
JOURNAL        Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
REFERENCE      4 (bases 1 to 178757)
AUTHORS        Parkway, St. Louis, MO 63108, USA
TITLE          Wilson,R.K.
JOURNAL        Submitted (22-MAY-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE      5 (bases 1 to 178757)
AUTHORS        Parkway, St. Louis, MO 63108, USA
TITLE          Wilson,R.
JOURNAL        Direct Submission
REFERENCE      Submitted (27-NOV-2003) Department of Genetics, Washington
AUTHORS        University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE          On May 22, 2003 this sequence version replaced gi:21105973.
JOURNAL        ----- Genome Center
COMMENT        Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
                ----- Summary Statistics
                -----
                Center project name: M_BB0443G20
                -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. 178757
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="16"

FEATURES             source
repeat_region        /map="16"
repeat_region        /clone="RP24-443G20"
repeat_region        /clone_lib="RPCI-24"
repeat_region        1123..1175
repeat_region        /rpt_family="B4"
repeat_region        2269..2522
repeat_region        /rpt_family="L1"
repeat_region        4511..4687
repeat_region        /rpt_family="CR1"
repeat_region        5027..5985
repeat_region        /rpt_family="L1"
repeat_region        6598..6703
repeat_region        /rpt_family="MER103"
repeat_region        9601..9658
repeat_region        /rpt_family="ERV1"
repeat_region        11560..11906
repeat_region        /rpt_family="MaLR"
repeat_region        12066..12931
repeat_region        /rpt_family="L1"
repeat_region        13001..13365
repeat_region        /rpt_family="MaLR"
repeat_region        13640..14068
repeat_region        /rpt_family="MaLR"
repeat_region        14376..14602
repeat_region        /rpt_family="MIR"
repeat_region        17949..18060
repeat_region        /rpt_family="B4"
repeat_region        18213..18359
repeat_region        /rpt_family="Alu"
repeat_region        19039..19112
repeat_region        /rpt_family="L2"
repeat_region        21794..21861
repeat_region        /rpt_family="L1"
repeat_region        21930..21985
repeat_region        /rpt_family="ERV1"
repeat_region        22019..22357
repeat_region        /rpt_family="L1"
repeat_region        25288..25415
repeat_region        /rpt_family="Alu"
repeat_region        25795..26039
repeat_region        /rpt_family="B4"
repeat_region        26754..27200
repeat_region        /rpt_family="ERV1"
repeat_region        27299..27440
repeat_region        /rpt_family="B4"
repeat_region        30863..31039
repeat_region        /rpt_family="MaLR"
repeat_region        31108..31319
repeat_region        /rpt_family="MaLR"
repeat_region        31433..31532
repeat_region        /rpt_family="Alu"
repeat_region        32345..32332
repeat_region        /rpt_family="B4"
repeat_region        32333..32460
repeat_region        /rpt_family="Alu"
repeat_region        32461..32522
repeat_region        /rpt_family="B4"
repeat_region        32753..32962
repeat_region        /rpt_family="L2"
repeat_region        34404..34484
repeat_region        /rpt_family="B4"
repeat_region        34432..34628
repeat_region        /rpt_family="Alu"
repeat_region        34676..34738
repeat_region        /rpt_family="Alu"
repeat_region        34870..34987
repeat_region        /rpt_family="B4"
repeat_region        37212..37345
repeat_region        /rpt_family="L2"
repeat_region        40253..40414
repeat_region        /rpt_family="L1"
repeat_region        40656..41180
repeat_region        /rpt_family="L1"

```


[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2004, 05:01:29 ; Search time 611 Seconds

(without alignments)
10825.605 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaacacatcc.....ccgttcagccgcgcagacaa 1557

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1557	100.0	1557	6	ABQ94220 FLO11 gen
2	1557	100.0	1800	6	ABQ94276 Fungal ge
3	1557	100.0	1833	6	ABQ94275 Fungal ge
4	391.6	25.2	867	7	ABZ51240 Aspergill
5	280.8	18.0	612	7	ABZ54133 Aspergill
6	59	3.8	236	3	ABZ57129 Pinus rad
7	51	3.3	5452	9	ADC86736 Human GPC
8	49.2	3.2	1856	4	ABL28791 Drosophil
9	49.2	3.2	3219	4	ABL28790 Drosophil
10	49.2	3.2	6530	2	AAQ51557 Lorixin
11	49.2	3.2	6530	2	AAZ22072 Nucleotid
12	49.2	3.2	6530	4	AAC68948 Mouse lor
13	48.6	3.1	3073	7	ADL51683 Human nuc
14	46.6	3.0	549	3	AAC93730 Cat flea
15	46.6	3.0	2000	7	ADA71938 Rice gene
16	46	3.0	110000	4	AAI99682_16
17	46	3.0	110000	4	AAI99683_16
18	45.2	2.9	492	2	AAT08579 Partial g
19	45.2	2.9	720	5	AAS05643 Mammalian
20	45.2	2.9	1476	2	AAT08578 Zinc fing
21	45	2.9	3133	9	ADC86738 Human GPC
22	44.6	2.9	3269	4	ABL26283 Drosophil
23	44.6	2.9	12920	4	ABL26282 Drosophil

24	44.4	2.9	4409	4	ABL17895	Ab117895 Drosophil
25	44.4	2.9	5021	4	ABL11783	Ab111783 Drosophil
26	44.4	2.9	7394	4	ABL17894	Ab117894 Drosophil
27	44.4	2.9	7614	4	ABL11782	Ab111782 Drosophil
28	44.4	2.9	10732	3	AAA10594	Aa10594 Gene enco
29	44.4	2.9	31814	7	AAD47150	Aad47150 Human Ras
30	43.8	2.8	2480	4	ABL25810	Ab125810 Drosophil
31	43.8	2.8	2874	4	ABL28677	Ab128677 Drosophil
32	43.8	2.8	5125	4	ABL28676	Ab128676 Drosophil
33	43.2	2.8	646	4	AAS26107	Aas26107 Human cDN
34	43.2	2.8	646	7	ABX73448	Abx73448 Human nov
35	43	2.8	336	2	AAV86698	Aav86698 EST clone
36	43	2.8	1506	4	ABL29105	Ab129105 Drosophil
37	43	2.8	1696	7	ACC83475	Acc83475 Mouse cho
38	43	2.8	2336	3	AAZ35052	Aaz35052 Mouse CNR
39	43	2.8	2651	3	AAZ35053	Aaz35053 Mouse CNR
40	43	2.8	3215	4	ABL29108	Ab129108 Drosophil
41	43	2.8	3506	4	ABL29104	Ab129104 Drosophil
42	43	2.8	3847	3	AAZ35044	Aaz35044 Mouse CNR
43	43	2.8	4841	4	AAK52955	Aak52955 Human pol
44	43	2.8	4880	4	AAK51971	Aak51971 Human pol
45	43	2.8	4898	7	ABZ79896	Abz79896 Human nuc

ALIGNMENTS

RESULT 1

ABQ94220

ID ABQ94220 standard; DNA; 1557 BP.

XX AC ABQ94220;

XX AC ABQ94220;

DT 22-OCT-2002 (first entry)

XX XX

DE FLO11 gene expression regulator An09 coding sequence.

XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At22; At24; At27; At32;
KW PC05; PC06; PC07; PC08; PC09; PC10; PC18; PC24; PC25; PC33; PC34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; PC1000; PC1001;
KW lovF; lovE; lovastatin; PC804; acvA; penicillin; antifungal; gene; ds.

XX OS Unidentified.

XX WO200257456-A2.

XX PD 25-JUL-2002.

XX PF 24-DEC-2001; 2001WO-US049911.

XX PR 22-DEC-2000; 2000US-0257431P.

XX PA (MICR-) MICROBIA.

XX PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
Holtzman D, Sherman A;

XX DR WPI; 2002-627368/67.

XX DR P-PSDB; ABP63087.

XX PT New isolated or recombinant gene, or purified protein, useful in
regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
production of enzymes, secondary metabolites or other commercially and
medically useful products.

XX PS Claim 1; Page 36-37; 71pp; English.

XX CC The present invention relates to novel fungal gene expression regulators
(ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,

CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc10, Pc18, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are lovE gene
CC expression regulators. lovF and lovE are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs
XX
SQ Sequence 1557 BP; 329 A; 533 C; 385 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 1557; DB 6; Length 1557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCTTAGAAACCATCCCTCTCGGCTCCGCTACCAAGTCTGCCTCAAGATCGGCG 60
Db
QY 61 CTTCTTCTTCTGCTCCATCTCGAGCATGCGCAATGCTCTGAGTACAGGATCGGCG 60
Db
QY 121 TACCCAGTCTCTCAGCGGACACCTGCTCTCTGCAACCCCATATAGCCAGTCTGCGC 180
Db
QY 121 TACCCAGTCTCTCAGCGGACACCTGCTCTCTGCAACCCCATATAGCCAGTCTGCGC 180
QY 181 GCTCTCTACTCTGATGGGGCAGCGCGCTGATCGGCTGATCTGAAACAGGTACCCCGCA 240
Db
QY 241 TCAAGTCAAGATGTTAGCGGCTTCTGCTGCGCGGATAATGCCCACTACCGTGGCG 300
Db
QY 301 AGCTTGCCTCCGACATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
Db
QY 301 AGCTTGCCTCCGACATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
QY 361 CAATCGCGCACTATCT 420
Db
QY 421 TACCCGCGAGCAATGGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Db
QY 481 CTTCTCTAGGAGCATTCAGTTATTTCGACGGAAGAGCTTGGACCCCGCTTGTCT 540
Db
QY 541 GCGAATGGTGTGCGCGCTATCCCGGAGGAGTCTCCCGGAGGAGTCTCCCGGAGGAGTCT 600
Db
QY 601 CAGGGCGAGCGGATTCCT 660
Db
QY 661 GTTAAAGGACCGGAGCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 720
Db
QY 721 TGCCTCGAAGTGTAAACAGACTTATCTTCTCATGCAAGCATCTCAAGGGCCATCTGCTACGC 780

Db 721 TGCCCGAAGTGTAAACAGACTTATCTTCTCATGCCAAGCATCTCAAGCGCATCTGTACGC 780
QY 781 CACTGTGTGACCGCCCGGTACATGTGTCTTTTTCGAAAGACACCTTCTCTCGCAGTAT 840
Db 781 CACTGTGTGACCGCCCGGTACATGTGTCTTTTTCGAAAGACACCTTCTCTCGCAGTAT 840
QY 841 ATCTGAAACGTCTATTTTCCAAAATGCTCAATCAGGCGTGTAAACCCACCGGACACG 900
Db 841 ATCTGAAACGTCTATTTTCCAAAATGCTCAATCAGGCGTGTAAACCCACCGGACACG 900
QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGGCAATCCT 960
Db 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGGCAATCCT 960
QY 961 GTAAAACTGTCCAGGATGAAGTCAAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020
Db 961 GTAAAACTGTCCAGGATGAAGTCAAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020
QY 1021 ACGACTTACCGGAGGAGCGCTCAACGGCAATGCACTAGCTCCGCGCCCGGCGGAGTAC 1080
Db 1021 ACGACTTACCGGAGGAGCGCTCAACGGCAATGCACTAGCTCCGCGCCCGGCGGAGTAC 1080
QY 1081 CGCGATCAACAGACTATGGGCTTCCCAATGTCATCCGTCAACGGGATGGCGGCTGTGTCAG 1140
Db 1081 CGCGATCAACAGACTATGGGCTTCCCAATGTCATCCGTCAACGGGATGGCGGCTGTGTCAG 1140
QY 1141 CCTGAAAGCGGCTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCCAAGGCTCC 1200
Db 1141 CCTGAAAGCGGCTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCCAAGGCTCC 1200
QY 1201 AAGCAGAGCGGCTATCTGTCAGCGGCTGTCACCTCTCTGCGCCACCATTTGAATATT 1260
Db 1201 AAGCAGAGCGGCTATCTGTCAGCGGCTGTCACCTCTCTGCGCCACCATTTGAATATT 1260
QY 1261 GACCGAATCATCGAGCAGGTAATAACACCGGCTTGTCAAGACCCCAAGCGCTGTGATG 1320
Db 1261 GACCGAATCATCGAGCAGGTAATAACACCGGCTTGTCAAGACCCCAAGCGCTGTGATG 1320
QY 1321 CCAGACATCCCGGCGACCCCGGTGAGCTTGAAGTGTGAGCTTATGTTCACCTCAAGCT 1380
Db 1321 CCAGACATCCCGGCGACCCCGGTGAGCTTGAAGTGTGAGCTTATGTTCACCTCAAGCT 1380
QY 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGTCAGGCTCAAGAGCCCATCCAGCTCAT 1440
Db 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGTCAGGCTCAAGAGCCCATCCAGCTCAT 1440
QY 1441 GTCGAGACCGGAGAAAGTATTACCCCACTACCGCTGCTCAAGAGAGTGAATGAAC 1500
Db 1441 GTCGAGACCGGAGAAAGTATTACCCCACTACCGCTGCTCAAGAGAGTGAATGAAC 1500
QY 1501 GGTCTCTATCTGGCTTGCATATGAGTGGGAGGACCGCTTACGCGCGGAGACAA 1557
Db 1501 GGTCTCTATCTGGCTTGCATATGAGTGGGAGGACCGCTTACGCGCGGAGACAA 1557

RESULT 2
ABQ94276
ID ABQ94276 standard; DNA; 1800 BP.
XX
AC ABQ94276;
XX
DT 22-OCT-2002 (first entry)
XX
DE Fungal gene expression regulator coding sequence SEQ ID 118.
XX
KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc24; Pc25; Pc33; Pc34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; Pc1000; Pc1001;
KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.

XX OS Unidentified.
XX PN WO200257456-A2.
XX PD 25-JUL-2002.
XX PF 24-DEC-2001; 2001WO-US049911.
XX PR 22-DEC-2000; 2000US-0257431P.
XX PA (MTCR-) MICROBIA.
XX PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
XX PI Holtzman D, Sherman A;
XX DR WPI; 2002-627368/67.
XX PT New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products.
XX PS Claim 67; Page 66-67; 71pp; English.
XX CC The present invention relates to novel fungal gene expression regulators
CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are lovE gene
CC expression regulators. lovF and lovE are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs
XX SQ Sequence 1800 BP; 372 A; 606 C; 463 G; 359 T; 0 U; 0 Other;
Query Match 100.0%; Score 1557; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGATCCTAGAAACCATCCCTTCGCGCTCCGCTACCGTCTACCGTCTCGCTCAAGGATCGGCG 60
DB 244 ATGGATCCTAGAAACCATCCCTTCGCGCTCCGCTACCGTCTCGCTCAAGGATCGGCG 303
QY 61 CCTCTTCTTCTGCTCCCATCTCGAGCATGCGCATGCGCTCAGTACAGATGCGACCTCAG 120
DB 304 CCTCTTCTTCTGCTCCCATCTCGAGCATGCGCATGCGCTCAGTACAGATGCGACCTCAG 363
QY 121 TACCCAGTCTCTAGCGGCACACCCCTGCTCTCTGCAACCCCATCATAGCAGTCCGCC 180
DB 364 TACCCAGTCTCTAGCGGCACACCCCTGCTCTCTGCAACCCCATCATAGCAGTCCGCC 423
QY 181 GCTCCTCAGTCTAGCGGCACACCCCTGCTCTCTGCGCGGATTAATGCCCCACACTACCGTGGC 240
DB 424 GCTCCTCAGTCTAGCGGCACACCCCTGCTCTCTGCGCGGATTAATGCCCCACACTACCGTGGC 483
QY 241 TCAAGTACAGATGTTTACGCGCTCTCTGCTGCGCGGATTAATGCCCCACACTACCGTGGC 300
DB 484 TCAAGTACAGATGTTTACGCGCTCTCTGCTGCGCGGATTAATGCCCCACACTACCGTGGC 543
QY 301 AGCTTGCTCCGACATCTTTTCCCTTCTCATCCCAATCCGAGGCGCGACAGGCGCAG 360

DB 544 AGCTTGCTCCGACATCTTTTCCCTTCTCATCCCAATCCGAGGCGCGACAGGCGCAG 603
QY 361 CAATCGCGGCACTATCCTCCTCATAGCGTGTCTCCGCCGCTTCCAGCGCTCAGTCG 420
DB 604 CAATCGCGGCACTATCCTCCTCATAGCGTGTCTCCGCCGCTTCCAGCGCTCAGTCG 663
QY 421 TACCGGAGCAATTTGCGCGGCGCTTCCCGGAGCGCTCCGCTGTGTGACTTCAACAATGGA 480
DB 664 TACCGGAGCAATTTGCGCGGCGCTTCCCGGAGCGCTCCGCTGTGTGACTTCAACAATGGA 723
QY 481 CTTCCTTCAGAGCATTCAGTTATTCGAGCGGAAAGCCTCAAGTTGGGACCCCGTGTCT 540
DB 724 CTTCCTTCAGAGCATTCAGTTATTCGAGCGGAAAGCCTCAAGTTGGGACCCCGTGTCT 783
QY 541 GCGAATGCTGCTGCGCGGTATCCCGGAGAGACTCCCGGAGAGCTCCCGGAGAGCTTGTGTCT 600
DB 784 GCGAATGCTGCTGCGCGGTATCCCGGAGAGACTCCCGGAGAGCTTGTGTGTGTGTCT 843
QY 601 CAGGGGCGAGCGCGGTATCCTTCCGAGTGTTCGCGGAGCGCAACTCCGCTCACAAATGGT 660
DB 844 CAGGGGCGAGCGCGGTATCCTTCCGAGTGTTCGCGGAGCGCAACTCCGCTCACAAATGGT 903
QY 661 GTTAAACCGGACCGGCAAGAACATATATATCCCGGCGCAAGGATCCGATGGAAAGTTCCCT 720
DB 904 GTTAAACCGGACCGGCAAGAACATATATATCCCGGCGCAAGGATCCGATGGAAAGTTCCCT 963
QY 721 TGGCCGAACTGTACAGAGCTTATCTTCATGCCAAGCATCTCAAGCGCCATCTCTGCTACGC 780
DB 964 TGGCCGAACTGTACAGAGCTTATCTTCATGCCAAGCATCTCAAGCGCCATCTCTGCTACGC 1023
QY 781 CACACTGCTGACCGCGGTACATGTGTGTCTTTTCAAGACACCTTCTCTCGCAGTGTAT 840
DB 1024 CACACTGCTGACCGCGGTACATGTGTGTCTTTTCAAGACACCTTCTCTCGCAGTGTAT 1083
QY 841 ATCTGAAACGTCTATTTCCAAAATGCTCAATCAGGCGTGTGTAAACCCACCGGAGCAACG 900
DB 1084 ATCTGAAACGTCTATTTCCAAAATGCTCAATCAGGCGTGTGTAAACCCACCGGAGCAACG 1143
QY 901 CACTTGTGCGACCCCATGCGCATGTGAAGAGTGTCCAAAGAGGCTGCGGCGCAATCTCT 960
DB 1144 CACTTGTGCGACCCCATGCGCATGTGAAGAGGTTCCAAAGAGGCTGCGGCGCAATCTCT 1203
QY 961 GTAAAACCTCTCCAGGATGAAGTACAGTACCGTCCCGCTCCCAATGGCATCCCGGCG 1020
DB 1204 GTAAAACCTCTCCAGGATGAAGTACAGTACCGTCCCGCTCCCAATGGCATCCCGGCG 1263
QY 1021 AGCACTTACGGCGAGGAGCGGTCAACGGCAATGGACTAGTCCGCGCGCGGCGGAGTAC 1080
DB 1264 AGCACTTACGGCGAGGAGCGGTCAACGGCAATGGACTAGTCCGCGCGCGGCGGAGTAC 1323
QY 1081 GCGGATCACCAGACTATGGGCTTCCCAATGTCTATCCGTCAACGGGATGGGCGGTGTCTAG 1140
DB 1324 GCGGATCACCAGACTATGGGCTTCCCAATGTCTATCCGTCAACGGGATGGGCGGTGTCTAG 1383
QY 1141 CCTGAAGACGCGTTTCCCGGCGCGCGCGCATCAAGAGGCGCCCTTGGGCCACCAAGCTCCC 1200
DB 1384 CCTGAAGACGCGTTTCCCGGCGCGCGCGCATCAAGAGGCGCCCTTGGGCCACCAAGCTCCC 1443
QY 1201 AAGCAGAGCGCGTATCTCTGTCAGCGCGGTGTGACCCCTTCTGGCCACCGAGTTGAATATT 1260
DB 1444 AAGCAGAGCGCGTATCTCTGTCAGCGCGGTGTGACCCCTTCTGGCCACCGAGTTGAATATT 1503
QY 1261 GACCGAAACATCAGCAGAGGTAAACAAACCGTTGTGTTCAAGACCCCAAGCCCTGTGTATG 1320
DB 1504 GACCGAAACATCAGCAGAGGTAAACAAACCGTTGTGTTCAAGACCCCAAGCCCTGTGTATG 1563
QY 1321 CCAGGACATCCCGGCGCACCCCGGTGTGACTGTGAGCGTCTATGTTTCCAACTCAAGCT 1380
DB 1564 CCAGGACATCCCGGCGCACCCCGGTGTGACTGTGAGCGTCTATGTTTCCAACTCAAGCT 1623
QY 1381 CCGAGGCGGTACATGTTCTCCCAAGTCTATGCTCTGTGTGTGTCAAGAGCCATCCAGCTCAT 1440

Db 1624 CCGAGGGCTACATGTTCTCCAGTCTATGCTCGTGGTGAAGAGCCCATCCACGCTCAT 1683
QY 1441 GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGCTGGTCAAGAGAGTGAATGAAC 1500
Db 1684 GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGCTGGTCAAGAGAGTGAATGAAC 1743
QY 1501 GGTCTCTATCTGGCTTCGACTATGATGTCGAGCGGACCGCTTCAGCCCGCCAGACAA 1557
Db 1744 GGTCTCTATCTGGCTTCGACTATGATGTCGAGCGGACCGCTTCAGCCCGCCAGACAA 1800

RESULT 3
ABQ94275
ID ABQ94275 standard; DNA; 1833 BP.
AC ABQ94275;
XX
DT
XX
22-OCT-2002 (first entry)
XX
DE Fungal gene expression regulator coding sequence SEQ ID 117.
XX
KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
KW lovE; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.
XX
OS Unidentified.
XX
PN WO200257456-A2.
XX
XX 25-JUL-2002.
XX
XX 24-DEC-2001; 2001WO-US049911.
XX
XX 22-DEC-2000; 2000US-0257431P.
XX
XX (MICR-) MICROBIA.
XX
XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
XX Holzman D, Sherman A;
XX WPI; 2002-627368/67.
XX
XX New isolated or recombinant gene, or purified protein, useful in
XX PT regulating fungal gene expression of FLO11, lovE, lovE or acvA for the
XX PT production of enzymes, secondary metabolites or other commercially and
XX PT medically useful products.
XX
XX Claim 67; Page 66; 71pp; English.
XX
XX The present invention relates to novel fungal gene expression regulators
XX CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
XX CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
XX CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
XX CC Pc08, Pc09, Pc10, Pc18, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
XX CC expression regulators. FLO11 is required for fungal invasion and its
XX CC expression is believed to be regulated by factors that also modulate
XX CC secondary metabolite production. At279, At286, At291, At320, At322,
XX CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
XX CC lovE gene expression regulators, and At501 and At574 are lovE gene
XX CC expression regulators. lovE and lovE are believed to be involved in the
XX CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
XX CC expression regulator. acvA is involved in the production of the secondary
XX CC metabolite penicillin. The fungal gene expression regulators and their
XX CC coding sequences are useful in regulating or manipulating the expression
XX CC of fungal genes that are involved in the production of enzymes, secondary
XX CC metabolites and other commercially and medically useful products, in
XX CC order to achieve maximum benefit. The genes may also be used to identify
XX CC genes relevant to fungal invasion which may act as targets for the
XX CC development of antifungal drugs

XX Sequence 1833 BP; 386 A; 613 C; 469 G; 365 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 1557; DB 6; Length 1833;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGATCTCTAGAAACCATCCCTCTCGGCTCGCTACAGTCTGCTCAAGGATCGGG 60
Db 277 ATGGATCTCTAGAAACCATCCCTCTCGGCTCGCTACAGTCTGCTCAAGGATCGGG 336
QY 61 CCTCTTCTCTGCTCCCATCTCGAGCATGCAATGCTCACTAGTACACAGTACAGCTCAG 120
Db 337 CCTCTTCTCTGCTCCCATCTCGAGCATGCAATGCTCACTAGTACACAGTACAGCTCAG 396
QY 121 TACCCAGTCTCTCAGCGGCACACCTGCTCTCTGCAACCCCATCATAGCCAGTCCGCC 180
Db 397 TACCCAGTCTCTCAGCGGCACACCTGCTCTCTGCAACCCCATCATAGCCAGTCCGCC 456
QY 181 GCTCTCTCATCTGATACATGCGGCGAGCGGCTACCGGCTGATCTGAACAGGTACCCCGCA 240
Db 457 GCTCTCTCATCTGATACATGCGGCGAGCGGCTACCGGCTGATCTGAACAGGTACCCCGCA 516
QY 241 TCAAGTCAAGATGTTTACGCGTCTTCTGCTGCGCGATAATGCTCCCACTACCTGCGC 300
Db 517 TCAAGTCAAGATGTTTACGCGTCTTCTGCTGCGCGATAATGCTCCCACTACCTGCGC 576
QY 301 AGCTTGCTCTCGACATCTTCTCTTCTCATCCCAATCGCAGGCGCAGCAGGCGCAG 360
Db 577 AGCTTGCTCTCGACATCTTCTCTTCTCATCCCAATCGCAGGCGCAGCAGGCGCAG 636
QY 361 CAATCGCGCACTATCTCTCTCTCATAGCGTCTCCGCGCGCTTCCAGCGCTCAGTGC 420
Db 637 CAATCGCGCACTATCTCTCTCTCATAGCGTCTCCGCGCGCTTCCAGCGCTCAGTGC 696
QY 421 TACCCGAGCAATTTGCGCGCGGCGCTCCCGGAGCGCTGCTGCTGACTTCAACATGA 480
Db 697 TACCCGAGCAATTTGCGCGCGGCGCTCCCGGAGCGCTGCTGCTGACTTCAACATGA 756
QY 481 CTTCCTTCAGGAGCATTCAGTTATTTCGAGCGAAAGCTCAAGTTGGACCCCGTGTCT 540
Db 757 CTTCCTTCAGGAGCATTCAGTTATTTCGAGCGAAAGCTCAAGTTGGACCCCGTGTCT 816
QY 541 GCGAATGCTGCTGCGCGCTATCCCGGAAAGACTCCCCCGGAAACCCAGTTTCTGTTCT 600
Db 817 GCGAATGCTGCTGCGCGCTATCCCGGAAAGACTCCCCCGGAAACCCAGTTTCTGTTCT 876
QY 601 CAGGGCGACGCGTATCTTCCGAGTCTTCCGGACGCGCAACTCCGCTCACAATGCT 660
Db 877 CAGGGCGACGCGTATCTTCCGAGTCTTCCGGACGCGCAACTCCGCTCACAATGCT 936
QY 661 GTTAAACGCGCACCGCAAGAACACTACTATCCCGGCAAGGATGCGGATGGAAGTCCCT 720
Db 937 GTTAAACGCGCACCGCAAGAACACTACTATCCCGGCAAGGATGCGGATGGAAGTCCCT 996
QY 721 TGCCGGAATGTAACAAGCATTTCTTATGCAAGCATCTCAAGCGGCATCTGCTACGC 780
Db 997 TGCCGGAATGTAACAAGCATTTCTTATGCAAGCATCTCAAGCGGCATCTGCTACGC 1056
QY 781 CACACTGCTGACCGCGCTGATGCTGTTCTTTTTCGAAGACACCTTCTCTGCAAGTGT 840
Db 1057 CACACTGCTGACCGCGCTGATGCTGTTCTTTTTCGAAGACACCTTCTCTGCAAGTGT 1116
QY 841 ATCTCTGAAACGCTCATTTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGACCAAG 900
Db 1117 ATCTCTGAAACGCTCATTTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGACCAAG 1176
QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTTCCTCAACAGAGGCTGCGGCGAATCT 960
Db 1177 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTTCCTCAACAGAGGCTGCGGCGAATCT 1236
QY 961 GTAAACCTGTCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020

Db 1237 GTAAACCTGTCCAGGATGAAGTCAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGGC 1296
QY 1021 ACAGCTTACCGCGAGGGAGCGGTCAACGGCAATGGACTAGTCTCGGCGCCGCCAGGGTAC 1080
Db 1297 ACAGCTTACCGCGAGGGAGCGGTCAACGGCAATGGACTAGTCTCGGCGCCGCCAGGGTAC 1356
QY 1081 GCGGATCACACAGATATGGGCTTCCCAATGTCAATCGTCAACGGGATGGGCGGTGGTCAAG 1140
Db 1357 GCGGATCACACAGATATGGGCTTCCCAATGTCAATCGTCAACGGGATGGGCGGTGGTCAAG 1416
QY 1141 CCTGAAGACGCTTCCCGCGCGCGCGCGCATCAAGAGAGCCCTTGGCCCAAGAGTCCC 1200
Db 1417 CCTGAAGACGCTTCCCGCGCGCGCGCGCATCAAGAGAGCCCTTGGCCCAAGAGTCCC 1476
QY 1201 AAGCAGAGCCGCTATCTCGTCAGCGCGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1260
Db 1477 AAGCAGAGCCGCTATCTCGTCAGCGCGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1536
QY 1261 GACCGAAACATCGAGCAGGTAAAAACAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1320
Db 1537 GACCGAAACATCGAGCAGGTAAAAACAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1596
QY 1321 CCAGGACATCCCGCCCAACCCCGGTGAGCTTGACCTGGAGCTGATGTTCCAACTCAAGCT 1380
Db 1597 CCAGGACATCCCGCCCAACCCCGGTGAGCTTGACCTGGAGCTGATGTTCCAACTCAAGCT 1656
QY 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1657 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1716
QY 1441 GTCGAGACCGGAGCAAGTATTACCCACCACTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 1717 GTCGAGACCGGAGCAAGTATTACCCACCACTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1776
QY 1501 GGTCTCTATCTGGTTCGACTATGAGTGGCGAGCGCACCGTTCAGCCCGCCAGCAA 1557
Db 1777 GGTCTCTATCTGGTTCGACTATGAGTGGCGAGCGCACCGTTCAGCCCGCCAGCAA 1833

RESULT 4
ABZ51240
ID ABZ51240 standard; cDNA; 867 BP.
XX
AC ABZ51240;
XX
DT 28-MAR-2003 (first entry)
XX
DE Aspergillus oryzae polynucleotide SEQ ID NO 353.
XX
KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW expressed sequence tag; gene; ss.
XX
OS Aspergillus oryzae.
XX
FN WO200279476-A1.
XX
PD 10-OCT-2002.
XX
PF 22-MAR-2002; 2002WO-IB000890.
XX
PR 30-MAR-2001; 2001JP-00098371.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (NARE-) NAT RES INST BREWING.
PA (NORQ) NAT FOOD RES INST MIN AGRTC.
XX
PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe X;
XX
DR WPI; 2003-046817/04.
XX
PT Detection of expression of specific *Aspergillus* genes for monitoring the
PT fermentation and growth conditions of the fungus, using DNA probes.

XX Claim 1; SEQ ID NO 353; 48bp + Sequence Listing; Japanese.
XX
CC The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of eutrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridising
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of *Aspergillus oryzae* which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 867 BP; 188 A; 257 C; 234 G; 182 T; 0 U; 6 Other;
Query Match 25.2%; Score 391.6; DB 7; Length 867;
Best Local Similarity 71.6%; Pred. No. 1e-103;
Matches 591; Conservative 1; Mismatches 206; Indels 27; Gaps 5;
QY 751 GCCAAGCATCTCAAGCCCATCTGCTACGCCAATCGGTGACCCCGGTACATGTGTGT 810
Db 31 GGCAAAACATCTGAAGCGCCATCTGCTCASACATACCGGTGACCCGCCATACATGTGTGTN 90
QY 811 CTTTGGCAAGACACCTTCTCTCGCAGTGATATCTGAAACGTCAATTTCCAAAATGCTCA 870
Db 91 CTGTGCAAGGATATCTTCTCCGCTAGTGATATCTGAAAGGCCACTTCCAGAAGTGTTCG 150
QY 871 ATCAGCGGTGTGTAACCCCAACCGGAGCAACGCACTTGTGCGACCCCAATGCGCATGTGAAG 930
Db 151 TTACGACGCGGCAACCCGACGGGAGCGCCACTTGTGCGATCCCGAGGACATTTGAAG 210
QY 931 AGGTCCCAACAGCAGGTGCGGGAATCTGTGTAACCTGTGTAAGATGAAGTCAGTAGT 990
Db 211 AGGTCTCAA-----GCTGGCAATCTCTGTAACCGGTTCAGGATGAAGTCAGTAGT 261
QY 991 ACCGTCCCGCTCCCAATGGCATCCCGGCGAGCACTTACCGGAGGAGCGGTCAACGGC 1050
Db 262 ACCGTCTCTCCCGCCACTGCTTCCCGGTACGGCTTACGGCGAGGGGGCGGTGAACGGT 321
QY 1051 AATGGACTAGTCTCGGCGCGCGGCGAGGATACGCGGATCACCAGACTATGGGCTTCCCAATG 1110
Db 322 AACGGGTGCGCTTCAGGCGGACCTGGGTTCAOGGATCAGCAGCCTCTGGGCTATCCGATG 381
QY 1111 TCATCCGTCAACGGGATGGGCGGTGCTGAGCTGAAGACGGTTCCTCGGCGCGCGGCGG 1170
Db 382 TCGTCGGTCAATGGGATGAACCGGTGTCACCGGACGATGCGTTCCTCGGTGTCAGCCA 441
QY 1171 CATCAAGGAGCCCTTGGCCCAAGCTCCCAAGCAGAGCCCGTATCTGTCAGCGCGGT 1230
Db 442 CATCAGAGAGCCCTTGGTGGCGGCTCCCAAGCAGAACCCGTTATCTGGCGCACTTGGC 501
QY 1231 GCTGACCCCTTCTGGCCACCACTTGAATATTGACCCGAAACATCGAGAGGTAAACCAACCG 1290
Db 502 ACTGATGTAATCTCAGCAGCTGAGTGTTCACCGTCTTCTTATGAACAGGTAAGCCCGCG 561
QY 1291 GTTGTTCAGACCCCAAGCGCCCTGTGATGCGCAGGAC-----ATCCGCGCCACCCCGGTGAG 1347
Db 562 GTGTGCAAGACCCCTAAGCGCCCAAGTAATGCTGGGCGCCACTTCCCAATCACTTGGCGAA 621
QY 1348 CTTGACCTGGAGCTTATGTTCCACCTCAAGCTCCCGAGGCGTACA-----TGTTTC 1398
Db 622 ATTGACTGGACTTCTGATGTTTCAGCCCGGAGCGAATGATGTTATACATCAACCCCGTTTC 681
QY 1399 TCCGAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1458
Db 682 CCCCAATCCATGGCTGTGCGCAGGAGCGGATCCATGCTCACGTCGATACCGACGGA 741
QY 1459 TATTACCCCAACCACTACCGTGTGT---CAAGAGAGTGAATGAACGGTCTCTATCTGCT 1515
Db 742 TTCTATCTCTACCAACACCGGTGTGTCAGGAGGTGCGATGATGTTGTTTACTTGGCT 801

CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERFs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
SO Sequence 236 Bp: 57 A: 82 C: 61 G: 36 T: 0 U: 0 Other:

	Query Match	3.8%;	Score 59;	DB 3;	Length 236;
	Best Local Similarity	61.3%;	Pred. No. 8.8e-07;		
	Matches	95;	Conservative	0;	Mismatches 60;
				Indels	0;
				Gaps	0;
QY	709	GGAAAGTTCCCTTGCCCGAACTGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGC	768		
Db	7	GGAGACCAGCTTGCCCGCTCTGCCAAGGACATTTTACTGCCCGCAACATGTGCGAGA	66		
QY	769	CATCTGCTAGCCACACTGTGTGACGGCCGCTACATGTGTCTTTTTCAAAGACACACCTTC	828		
Db	67	CACATGGCTCCACACACGGCGGACGGCCCGTACAAGTGCTCCATCTGCACCGACTCGTTT	126		
QY	829	TCTCGAGTGGATATCTCTGAACGTCATTCTCAAAA	863		
Db	127	GGCGGAGCGACCTCTCTGAAGCGACATCAGAAGAA	161		

RESULT 7	
ADC86736/c	
ID	ADC86736 standard; DNA; 5452 BP.
XX	
AC	ADC86736;
XX	
DT	01-JAN-2004 (first entry)
XX	
DE	Human GPCR gene SEQ ID NO:1189.
XX	
KW	ds; gene; human; GPCR;
KW	guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX	
OS	Homo sapiens.

PN	EP1270724-A2.
XX	
PD	02-JAN-2003.
XX	
PF	18-JUN-2002; 2002EP-00013517.
XX	
PR	18-JUN-2001; 2001JP-00246789.
XX	
PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA	(ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX	
PI	Suwa M, Asai K, Akiyama Y, Aburatani H;
XX	
DR	WPI; 2003-315783/31.
DR	P-PSDB; ADC86737.
XX	
PT	New polynucleotide, useful for preparing a composition for treating a
PT	patient in need of increased or suppressed activity or expression of the
PT	guanine triphosphate-binding protein coupled receptor.
PT	

xx Claim 1; SEQ ID NO 1189; 28pp; English.

xx

xx The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC8548-ADC87616 encode GPCR's of the
CC invention.

xx Sequence 5452 BP; 796 A; 75 C; 1826 G; 79 T; 0 U; 2676 Other;

xx

[illegible]

RESULT 8	
ABL28791	
ID	ABL28791 standard; DNA; 1856 BP.
XX	
AC	ABL28791;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 37846.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
PR	11-JUL-2000; 2000US-00614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 o
PT	genes from Drosophila and for elucidating cell signaling and cel
PT	interactions.
XX	
PS	Claim 1; SEQ ID NO 37846; 21pp + Sequence Listing; English.
XX	

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (AB557737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published pct sequences

Query Match 3.2%; Score 49.2; DB 4; Length 1856;
Best Local Similarity 59.2%; Pred. No. 0.0016;
Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy	714	GTTCCTTGCCCGAACTGTAA	CAAGACTTATCTTTCATGCCAAGCATCTCAAGCGCCATCT	773
Db	1207	GTACTCTGCAAAATGTGCAGCAAGAGCTTTGTGCAGTCCCAAGACCTTAAGATCCATAT		1266

Qy	774	GCTACGCACACATGGTGTACCGCCCGTACATGTGTCTTTGCAAGACACACTTCTCTCG	833
Db	1267	GGGGCGCCACACGGCGAGCGACCATATCAGTGTGGAGTGTGGGCGAGAGCTTTGTGTG	1326

Qy		834	CAGTGAATCTCCTGAAACGTCA	855
Dβ		1327	CGTTTCGCACCCTGAATATCCAT	1348

RESULT 9
ABL28790/c
ID ABL28790 standard; DNA; 3219 BP.

AC	ABL28730;
AC	
XX	
DT	26-MAR-2002
DT	(first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;

XX
OS
XX

Drosophila melanogaster.

XX
XX

XX
PD
XX
27-SEP-2001.

XX
23-MAR-2000; 2000US-0191637P.
PR
11-JUN-2000; 2000US-00614150.
PR

PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;

DR	WPI; 2001-656860/75.
XX	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more

PT interactions.

XX

PS Claim 1; SEQ ID NO 37843; 2lpp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and

insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published/pct/sequences

AA
SQ Sequence 3219 BP; 842 A; 747 C; 731 G; 899 T; 0 U; 0 Other;
Query Match 3.2%; Score 49.2; DB 4; Length 3219;

Qy 714 GTTCGCTTGCCGAACTGTAAAGACTATCTTCATGCGAAGCATCTCAACGGCCATCT

960 GTACTCCTGCAAAATGTGCAGCAAGAGCTTTGTGCAGTCCAAACGACCTTAAGATCCATAT
960 GTACTCCTGCAAAATGTGCAGCAAGAGCTTTGTGCAGTCCAAACGACCTTAAGATCCATAT
774 GCTAGGCCCACTGTGTGACCGCCCGTACATGTGTCTTTTGCAAGACACCTTCTCTCG

Db 900 GCGGGCCACACGGGGAGCGACCATATCAGTGTGGAGTGTGCGCGAGAGCTTTGTGTG

Qy 834 CAGTGATATCCTGAAACGTCAAT 855

Db 840 CGGTTCGCACCTGAATATCCAT 819

AAQ51557/c
ID AAQ51557 standard; DNA; 6530 BP.
XX

XX	
DT	25-MAR-2003 (revised)
DT	17-MAY-1994 (first entry)

DE Loricrin gene.
XX
KW Loricrin; gene; constitutive; inducible; vector; 5' flanking region;

DNA cassette; linker; restriction site; induction; human; psoriasis;
 KW
 epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
 KW
 transforming growth factor beta; cytokine; skin; tracheal epithelia;
 KW

KW capsid protein; squamous epithelium; viral; oesophageal epithelia;
 KW vaginal epithelia; corneal epithelia; transduction; transplant; ss.
 XX

XX	Key	Location/Qualifiers
PH		
FT	misc_RNA	1..1540
+		/+...

```
FT /note="5' flanking region"
FT 1587. .2677
FT /*tag= b
FT /note="5' flanking region"
FT 1587. .2677
```

```
FT misc_RNA      4384. .6530
FT                /*tag= c
FT                /note= "3' flanking region"
FT
FT
FT
```

PN WO9322431-A1.
XX
PD 11-NOV-1993.
yy

PF	28-APR-1993;	93WO-US003993.
XX		
PR	30-APR-1992;	92US-00876286.
vv		

PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Roop DR, Rothnagel JA, Greenhalgh DA;
XX

DR	WPI; 1993-368790/46.
XX	
PT	Constitutive and inducible vectors based on loricrin or keratin K6 genes

PT treating psoriasis and skin cancer, etc.

Claim 23; Page 26-32; 68pp; English.

This sequence represents the loricrin gene and was used in the construction of the loricrin constitutive vector of the invention. The loricrin vector has a 5' flanking region of approx. 1.5 kb, an intron of approx. 1.1 kb and a 3' flanking region of approx. 1.5 kb. The 5' flanking region of the loricrin gene comprises a TATA box, cap site, first intron and intron/exon boundary, all in the correct order and position for expression of a nucleic acid cassette. The loricrin vector also comprises a 3' flanking region of the loricrin gene and a linker with a unique restriction site at the position of the start and stop codons, joining the two flanking sequences and also having a position for insertion of the DNA cassette (see also AAQ51558). Vectors such as this can be used for: (a) in vitro induction of human epidermal cells, and especially for stimulating healing of wounds, surgical incisions, and ulcers, (where the vector expresses a growth factor), (b) treating psoriasis, (the vector expressing transforming growth factor beta, a soluble cytokine or antisense RNA), (c) treating skin cancer, (the vector expressing antisense RNA of the B6 or E7 gene of human papilloma virus or normal p53 protein), (d) for vaccination (the vector expresses a viral capsid protein, especially of human papilloma virus), or (e) treating cancer of the squamous epithelium (of epidermal, oral, oesophageal, vaginal, tracheal or corneal epithelia, the vector expresses antisense RNA). Alternatively human epithelial cells are transduced ex vivo, then transplanted. (Updated on 25-WAR-2003 to correct PN field.)

Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 U; 0 Other;

Query Match	3.2%;	Score 49.2;	DB 2;	Length 6530;
Best Local Similarity	46.1%;	Pred. No. 0.0029;		
Matches 165;	Conservative 0;	Mismatches 193;	Indels 0;	Gaps 0;
QY	96	GCCTCAGTACACGATGACGCTCAGTACCCAGTCTCTCAGCGGCACACCCCTGCTCTCTCT	155	
Db	3530	GCCTCCGTAGCTCTGGCACTGATCTGTTGGGAGCATCCGCGCGCGCCGCCACCGAGGA	3471	
QY	156	GCAACCCCATCATAGCCAGTGCCTCGTCTCTCTCAGTCTATGATGGGAGCGCCGCTACCG	215	
Db	3470	ACCACTCCGAGCTCTGTCGCGCCACCGAATAGCCGCGCGCGCGAGCTGGA	3411	
QY	216	GCTGTATCTGAACAGGTACCCGCGCATCAGTCAAGTATTACGCGTCTTCTGTGCGCC	275	
Db	3410	TCCACCCCACTGAGTAGCGCGCCACAGTACAGCGCGCGCTGTTCCACGCGCGCC	3351	
QY	276	GATTAATGCCCCACACTACCGTGGCAGCTTGCCCTCCGACATCTTCTTCTCATCCCA	335	
Db	3350	GGAATAGCCGCTCCGAGCTGGAGCCACCGCCGCCACCGAGTCTGCGCGCTCCGGA	3291	
QY	336	TCGCGAGCGCGAGGCACAGGCGCAGCAATCGCCCACTATCGCTCTCTCATAGCGTGT	395	
Db	3290	ACCGCGCGCAGCCACCGAGCTGCCGCTCCAGAGAGGCTCCGAGTAGGAGCGCC	3231	
QY	396	CCGCGCGCTTCAGCGCTTAGTCGTACCCGACGCAATTGCGCGCGCGCTCCCGG	453	
Db	3230	GCCTCTGTATCCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGAGAGCGCG	3173	

RESULT 11	
AAZ22072/c	
ID	AAZ22072 standard; DNA; 6530 BP.
XX	
XX	
AC	AAZ22072;
XX	
DT	25-NOV-1999 (first entry)
XX	
DE	Nucleotide sequence of the lorocrin gene.
XX	
XX	Vector; epidermal cell; lorocrin gene; wound;
KW	surgical incision; disease; psoriasis; cancer; alopecia;
KW	transgenic animal; ss.
XX	
XX	
OS	Mus sp.

XX	US95958764-A.	
PN		
XX		
XX	28-SEP-1999.	
PD		
XX		
XX	01-NOV-1993;	93US-00146930.
PF		
XX		
XX	30-APR-1992;	92US-00876286.
PR		
XX	29-OCT-1993;	93US-00145388.
XX		
XX	(BAYU) BAYLOR COLLEGE MEDICINE.	
XX		
PA		
XX		
PI	Greenhalgh DA, Rothnagel JA, Roop DR;	
XX		
XX	WPI; 1999-561044/47.	
DR		
XX		
XX	Gene therapy vectors for expression in the epidermis.	
PT		
XX		
XX	Example 1; Col 25-32; 34pp; English.	
XX		
PS		
CC	The specification describes a vector for the expression of a nucleic acid	
CC	sequence in an epidermal cell, where the nucleic acid sequence is not a	
CC	mammalian lorixin gene or a mammalian K6 keratin gene. The vector	
CC	comprises a 5' flanking region including the sequences required for its	
CC	expression; a 3' region from a mammalian lorixin or a mammalian K6	
CC	keratin gene which regulates expression (predominantly in an epidermal	
CC	cell) of the nucleic acid sequence with which it is associated; and a	
CC	linker connecting the 5' flanking region to the nucleic acid sequence,	
CC	where the linker comprises a position for inserting the nucleic acid	
CC	sequence and lacks the coding sequence of a gene with which it is	
CC	naturally associated. The vector is able to encode a hormone, a growth	
CC	factor, an enzyme, a clotting factor, an apolipoprotein, a receptor, a	
CC	drug or an antigen. The vectors can be used to treat wounds or surgical	
CC	incisions as well as diseases, such as psoriasis, cancer, alopecia and to	
CC	create transgenic animals for assessing human disease in an animal model.	
CC	The present sequence represents the lorixin gene	

xx	SQ	Sequence	6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 U; 0 Other;
		Query Match	3.2%; Score 49.2; DB 2; Length 6530;
		Best Local Similarity	46.1%; Pred. No. 0.0029;
		Matches 165; Conservative	0; Mismatches 193; Indels 0; Gaps 0;
Qy	96	GCCTCAGTACACGATGCAGCGCTCAGTACCACTCTCTCAGCGGCACACCTGCTCCTCTCTCT 155	
Db	3530	GCCTCGTAGCTTGCGCACTGTACTCTTGGAGCATCGCGCGCGCGCACACGGAGGA 3471	
Qy	156	GCAACCCCATATAGCCAGTCGCGCGCTCTCTCACTCGTACATGGGGCAGCGCGCTACCG 215	
Db	3470	ACCACTCCGCAGCTCTGTGCGCGCGCACCGGATAGCCGCGCGCGCGCAGCTGGA 3411	
Qy	216	GCCTGATCTGAACAGGTACCCCGCATCAAGTCAAGATGTTTACGCGTCTTCTGTCTGCGCC 275	
Db	3410	TCCACCGCCACTGAGTAGCGCGCGCACAGCTAGAGCGCGCGTGCTTCCACGCGCGCC 3351	
Qy	276	GATATGCCCCACACTACCGTGGGAGCTTGCTCCGACATCTTCTCTTCTCATCCAA 335	
Db	3350	GGAATAGCGCCCTCCGAGCTGGAGCCACCGCGCCACCAAGAGTACTTGCCTCCCGGA 3291	
Qy	336	TCCGCAAGCGCAGGCACAGCGGACGACNAATCGCCGCACTATCTTCTCTCTCATAGCTGCT 395	
Db	3290	ACCGCGCGCGACGCCACCGAGCTGCGCGCTCCAGGAGGAGCTTCGAGTAGGAGCGCC 3231	
Qy	396	CCGCGCGGCTTCAGCGCTCAGTCTGTAACCGCAGCCNAATTGCGCGCGGCGCTCCCCG 453	
Db	3230	GCCTCTGTATCCGCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGCGGAGCGCGCG 3173	

RESULT 12
AAC68948/C
ID AAC68948 standard; DNA; 6530 BP.
XX
AC AAC68948:

```
XX DT 26-FEB-2001 (first entry)
XX DE Mouse loricrin gene.
XX KW Loricrin; cytotretic; gene therapy; cell proliferation inhibition;
XX KW murine; skin cancer; skin ulcer; psoriasis; alopecia; da.
XX OS Mus sp.
XX FN US6143727-A.
XX PD 07-NOV-2000.
XX PF 05-JUN-1995; 95US-00458240.
XX PR 30-APR-1992; 92US-00876286.
XX PR 29-OCT-1993; 93US-00145388.
XX PR 01-NOV-1993; 93US-00145930.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX FI Roop DR, Greenhalgh DA, Rothnagel JA;
XX DR WPI; 2001-006441/01.
XX CC Gene therapy of skin cancer involving administering specific expression
XX FT vector, loricrin or K6 keratin expression vector within the epidermis or
XX FT epidermal cells.
XX PS Claim 8; Col 25-32; 35pp; English.
XX CC The present invention relates to a method for treating skin cancer. The
XX CC method comprises administering a loricrin or K6 keratin expression vector
XX CC comprising a nucleic acid sequence encoding a wild-type p53 protein at or
XX CC directly around the site of skin cancer cell, where the expression of the
XX CC p53 protein by the skin cancer cell results in inhibition of the cell's
XX CC proliferation. The present sequence is the murine loricrin gene. This
XX CC gene was used to construct the loricrin expression vector of the present
XX CC invention. The method of the present invention may also be used to treat
XX CC skin ulcers, psoriasis and alopecia
XX SQ Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 U; 0 Other;
Query Match 3.2%; Score 49.2; DB 4; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.0029;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY 96 GCCTCAGTACAGATGCGCCTCAGTACCCAGTCTCTCAGCGCGCACACACCTGCTCTCT 155
DB 3530 GCCTCGTAGCTGGCACTGATCTGTTGGAGATCCCGCGCGCCACCGGAGGA 3471
QY 156 GCAACCCCAATATAGCAGTGGCGGCTCTCTACTCTATCATGCGGCGCGCGGTACCG 215
DB 3470 ACCACCTCCGAGCTGTGCGCGCGCCACCGAATAGCCGCCGCGCGCGAGCTGGA 3411
QY 216 GCCTGATCTGAACAGGTACCCGCATCAAGTACGATGTTTACGCGTCTCTCTGCGCC 275
DB 3410 TCCACCGCCACCTAGTAGTGGCGCGCCACAGTAGAGCCGCCGCTCTCCACCGCGCC 3351
QY 276 GATAATGCCCCACACTACCGTGGCGAGTTGGCTCCGACATCTTCTCTTCTATCCAA 335
DB 3350 GGAATAGCCGCTCCGAGCTGGAGCCACCGCGCCACAGAGTACTTGGCGCTCCGGA 3291
QY 336 TCCGAGGCGCAGCAGCAGCGCAGCAATCCCGCACTTCTCTCTCTCTATGCGTGT 395
DB 3290 ACCGCGCGCGCAGCAGCAGCGAGTGGCGCTCCAGAGGAGCTCCGAGTAGAGCGCGC 3331
QY 396 CCGCGCGCGCTTCCAGCGCTCAGTGTGTTACCGCGAGCCAAATGGCGCGCGCTCCCGG 453
DB 3230 GCCTCTGTATCCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGAGGAGCGCG 3173
```

```
RESULT 13
AAD51683
ID AAD51683 standard; cDNA; 3073 BP.
XX AC AAD51683;
XX DT 16-APR-2003 (first entry)
XX DE Human nucleic acid associated protein (NAAP)-9 encoding cDNA.
XX KW Human; nucleic acid associated protein; NAAP; cancer; atherosclerosis;
XX KW cell proliferative disorder; neurological disorder; Huntington's disease;
XX KW epilepsy; stroke; immune disorder; acquired immune deficiency syndrome;
XX KW AIDS; inflammatory disorder; allergy; developmental disorder; infection;
XX KW hypothyroidism; cushing's syndrome; gene therapy; cytostatic; nootropic;
XX KW anticonvulsant; neuroprotective; cerebroprotective; thyromimetic; gene;
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1055..2242
XX FT /*tag= a
XX FT /product= "Human NAAP-9 protein"
XX FN WO200299115-A2.
XX PD 12-DEC-2002.
XX PF 31-MAY-2002; 2002WO-US017050.
XX PR 01-JUN-2001; 2001US-0295359P.
XX PR 08-JUN-2001; 2001US-0296878P.
XX PR 15-JUN-2001; 2001US-0297222P.
XX PR 15-JUN-2001; 2001US-0298615P.
XX PR 15-JUN-2001; 2001US-0298665P.
XX PR 15-JUN-2001; 2001US-0298693P.
XX PR 21-JUN-2001; 2001US-0300176P.
XX PR 19-APR-2002; 2002US-0373891P.
XX PA (INCY-) INCVTE GENOMICS INC.
XX PA (YUEH/) YUE H.
XX PI Yue H, Tang YT, Baughn MR, Becha SD, Warren BA, Walia NK;
XX PI Lal PG, Lee EA, Harfalla AJA, Richardson TW, Griffin JA, Emerling BM;
XX PI Ramkumar J, Yue H, Swarnakar A, Tran B, Li JX, Yao MG, Yang J;
XX PI Ison CH, Forsythe IJ, Honchell CD, Arvizu CS, Elliott VS, Lu Y;
XX PI Ding L, Luo W, Wang YE, Burford N, Borowsky ML, Nguyen DB;
XX PI Chinn AM, Kabie AE;
XX WPI; 2003-140626/13.
DR P-PSDB; AAE33769.
XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,
XX PT treating and preventing diseases or conditions associated with the
XX PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
XX PT infections.
XX PS Claim 94; Col 235-236; 257pp; English.
XX CC The present invention relates to human nucleic acid associated proteins
XX CC (NAAP) and polynucleotides encoding such proteins. NAAP sequences are
XX CC useful in diagnosing, treating and preventing diseases or conditions
XX CC associated with the decreased expression or overexpression of NAAP such
XX CC as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
XX CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
XX CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
XX CC disorders or infections. They are also used in gene therapy. The present
XX CC sequence is human NAAP-9 cDNA
XX SQ Sequence 3073 BP; 487 A; 1117 C; 970 G; 499 T; 0 U; 0 Other;
Query Match 3.1%; Score 48.6; DB 7; Length 3073;
```

Best Local Similarity 54.9%; Pred. No. 0.0031;
 Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
 QY 695 CCAAGGATGCGATGGAAGTTCCCTTCCCGAACTGTAAACAAGACTTATCTTCATGCCA 754
 Db 86 CCGGGGTGGGATGGCCGTTCCCTTCCCGAACTGTAAACAAGACTTATCTTCATGCCA 145
 QY 755 AGCATCTCAAGCGCATCTGTAGCCACATCTGTGTGACCGCCGATACATGTGTCTTT 814
 Db 146 GTCACCTTCCACACACAGATATCCACACGGCGAGAGCCCTTCTCTGTCCTGTGT 205
 QY 815 GCAAGACACCTTCTCTGCGATGATATCTTCCGAAAGCTATTTCCAAAAATGCTC 869
 Db 206 GCACCAAAGTTCAACCGCAGGAGAGTCTGAAGCCGACGCTGAAGACGCACTC 260

RESULT 14

AAC93730
 ID AAC93730 standard; cDNA; 549 BP.

XX AC AAC93730;

XX DT 19-FEB-2001 (first entry)

XX DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:225.

XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;

XX KW flea infestation; vaccine; antiparasitic; therapeutic target; diagnosis;
 detection; ss.

XX OS Ctenocephalides felis.

XX PN W0200061621-A2.

XX PD 19-OCT-2000.

XX PF 07-APR-2000; 2000WO-US009437.

XX PR 09-APR-1999; 99US-0128704P.

XX PA (HESK-) HESKA CORP.

XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;

XX DR WPI; 2000-656323/63.

XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 acids useful for the prevention, diagnosis and treatment of flea
 infestations.

XX PS Claim 26; Page 320; 964pp; English.

XX CC The invention relates to novel cat flea (*Ctenocephalides felis*) nucleic
 acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 or head and nerve cord (HNC) tissue. The invention also relates to the
 encoded proteins. The invention additionally encompasses expression
 constructs, recombinant viruses and recombinant cells comprising the
 nucleic acids of the invention, recombinant production of the proteins,
 antibodies against the proteins, a method of identifying inhibitors of
 the proteins, and compositions comprising the inhibitors for
 administration to an animal. The nucleic acids, and the proteins they
 encode may be used in the prevention, treatment and diagnosis of diseases
 associated with flea infestations. For example, the nucleic acids may be
 used to produce an HMT or HNC protein according to standard recombinant
 DNA methodology by inserting the nucleic acids into a host cell and
 culturing the cell to express the protein. The HMT and HNC nucleic acids
 may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
 and quantitate the presence of cat flea or other homologous nucleic acid
 sequences in samples. They may also be used to study the expression and
 function of the proteins and their role in metabolism. The HMT and HNC
 proteins may be used as antigens in the production of specific
 antibodies, and in assays to identify modulators (agonists and
 antagonists) of HMT and/or HNC protein expression and activity. The anti-

CC HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HMT cDNA of the invention
 XX
 SQ Sequence 549 BP; 170 A; 111 C; 114 G; 150 T; 0 U; 4 Other;

Query Match 3.0%; Score 46.6; DB 3; Length 549;
 Best Local Similarity 55.2%; Pred. No. 0.0054;
 Matches 91; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 697 AAGATCGCGATGGAAGTTCCCTTCCCGAACTGTAAACAAGACTTATCTTCATGCCAAG 756
 Db 219 AATCTCTGATGACATACACCGATGCGAACTGTGCAATAAGCGTTTAAATTGCAACA 278
 QY 757 CATCTCAAGCGCCATCTGTACGCCACACCTGGTGACCGCCCGTACATGTGTCTTTTGC 816
 Db 279 AATCTTAAAGCTCATATGTTATGCCACAGCGCGCAACGGCCCTATGCTGTGCCATATGC 338
 QY 817 AAGACACCTTCTCTCGCATGATATCTTGAAGCTCATTTCCAA 861
 Db 339 AGTAAAGCTTTTCCACGAATTGCCACCTTGAAACCGACACATGCAA 383

RESULT 15

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5263.

XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
 gene; ds.

XX OS *Oryza sativa*.

XX PN W02003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX DR WPI; 2003-175290/17.

XX PT Identifying at least one gene involved in plant resistance or response to
 pathogenic infection for conferring resistance or tolerance to a plant to
 bacterial, fungal or viral infection by determining or detecting plant
 gene expression.

XX PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2004, 05:22:55 ; Search time 3726 Seconds
(without alignments)
12478.641 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaaccatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pto:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gsa1:*
29: gb_gsa2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	380.4	24.4	442	9	AA785001
2	185	11.9	719	14	CF945290
3	171.6	11.0	776	12	BM077105
4	108.8	7.0	687	10	BF251517

5	82.8	5.3	782	28	AQ161501
6	82	5.3	585	28	AQ255041
7	56.2	3.6	807	13	BX464554
8	53.8	3.5	1448	28	CC220110
9	52.6	3.4	1201	9	AL515475
10	52.2	3.4	727	13	BX319543
11	52.2	3.4	1970	29	CG748837
12	51.4	3.3	1137	28	CC315110
13	51.2	3.3	1839	29	CG747711
14	50.6	3.2	567	28	AZ895651
15	50.6	3.2	1616	29	CG753270
16	50.2	3.2	916	29	CC678630
17	50.2	3.2	929	29	CG276224
18	50.2	3.2	939	29	CG287073
19	50.2	3.2	945	29	CC678642
20	50	3.2	1201	13	BX457568
21	49.6	3.2	1053	14	CK028030
22	49.6	3.2	1201	13	BX421672
23	49.4	3.2	802	29	CNS01170
24	49.2	3.2	786	28	AQ745021
25	49.2	3.2	897	29	AG060530
26	49.2	3.2	915	13	BQ952432
27	49.2	3.2	1262	12	BM560943
28	49.2	3.2	1287	10	BF256624
29	49	3.1	481	9	AA695157
30	49	3.1	759	13	BU314584
31	48.8	3.1	834	28	BZ051898
32	48.8	3.1	1101	29	CNS00008
33	48.8	3.1	1201	9	AL514195
34	48.6	3.1	1642	29	CG754720
35	48.6	3.1	458	10	BF063108
36	48.6	3.1	648	12	BI227246
37	48.6	3.1	1026	29	AG081192
38	48.6	3.1	1793	29	CG754612
39	48.4	3.1	732	29	BX149729
40	48.4	3.1	815	29	BX239144
41	48.4	3.1	987	29	CNS000418
42	48.4	3.1	1137	29	AG078502
43	48.2	3.1	786	28	BH018846
44	48.2	3.1	2152	28	CC283954
45	48	3.1	785	29	AG081505

ALIGNMENTS

RESULT 1

AA785001

LOCUS

DEFINITION

g4b03a1.r1 Aspergillus nidulans 24hr asexual developmental and

vegetative cDNA lambda zap library Emericella nidulans cDNA clone

g4b03a1 5', mRNA sequence.

ACCESSION

AA785001

VERSION

AA785001.1

KEYWORDS

EST.

SOURCE

Emericella nidulans

ORGANISM

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiiales; Trichocomaceae; Emericella.

REFERENCE

1 (bases 1 to 442)

AUTHORS

Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,

Prade,R. and Roe,B.

TITLE

An Aspergillus nidulans EST Database

JOURNAL

Unpublished (1998)

COMMENT

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

AA785001 442 bp mRNA linear EST 29-JUL-1998
g4b03a1.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Emericella nidulans cDNA clone
g4b03a1 5', mRNA sequence.

AA785001 GI:2845169

Emericella nidulans (anamorph: Aspergillus nidulans)

Emericella nidulans

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiiales; Trichocomaceae; Emericella.

1 (bases 1 to 442)

Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,

Prade,R. and Roe,B.

An Aspergillus nidulans EST Database

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: SK

High quality sequence stop: 429.

Location/Qualifiers

FEATURES

source
1. .442
/organism="Emericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone="g4b03a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual development and vegetative cDNA lambda zap library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

ORIGIN

Query Match 24.4%; Score 380.4; DB 9; Length 442;
Best Local Similarity 98.5%; Pred. No. 6.8e-79;
Matches 384; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 994 GTCCGCTCCGATCCGATCCGCGACGACTTACGGGAGGAGCGCGTCAACGGCAAT 1053
Db 53 GTCCGCTCCGATCCGCGACGACTTACGGGAGGAGCGCGTCAACGGCAAT 112

QY 1054 GGAAGTACGCTCCGCGCGCGGATACGGGATACGAGACTATGGCTTCCCAATGTCA 1113
Db 113 GGAAGTACGCTCCGCGCGCGGATACGGGATACGAGACTATGGCTTCCCAATGTCA 172

QY 1114 TCCGTCACGCGGATGGCGCGTGTGAGCTGAGAGCGCTTCCGCGCGCGCGCGCAT 1173
Db 173 TCCGTCACGCGGATGGCGCGTGTGAGCTGAGAGCGCTTCCGCGCGCGCGCGCAT 232

QY 1174 CAAAGGAGCGCTTGGCGACAAAGCTCCCAAGCAGAGCGCGTATCTGTGAGCGCGGTGT 1233
Db 233 CAAAGGAGCGCTTGGCGACAAAGCTCCCAAGCAGAGCGCGTATCTGTGAGCGCGGTGT 292

QY 1234 GACCTTCTGGCCACCAAGTGTGATATGACCGGAAACATCAGCAGTGAACAAACCGGTT 1293
Db 293 GACCTTCTGGCCACCAAGTGTGATATGACCGGAAACATCAGCAGTGAACAAACCGGTT 352

QY 1294 GTTCAAGACCCCAAGCGCGCTGTGATGCGCAGGACATCCCGCGCACCGCGTGTGAC 1353
Db 353 GTTCAAGACCCCAAGCGCGCTGTGATGCGCAGGACATCCCGCGCACCGCGTGTGAC 412

QY 1354 TGGAGCTCTATGTTCCAACTCAAGCTCC 1383
Db 413 TGGAGCTCTATGTTCCAACTCAAGCTCC 442

RESULT 2

CF945290 719 bp mRNA linear EST 19-NOV-2003
LOCUS TrEST-A3958 TrEST-A Hypocrea jecorina cDNA clone Tr-A3958 5', mRNA
DEFINITION sequence.
ACCESSION CF945290
VERSION CF945290.1 GI:38445810
KEYWORDS EST.

SOURCE

Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE

1 (bases 1 to 719)
Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,
Ferreira, J.R., Jr., Abrahao-Neto, J., Farah, J.P.S., and El-Dorri, H.
Elucidation of the metabolic fate of glucose in the filamentous
fungus Trichoderma reesei using expressed sequence tag (EST)

AUTHORS

analysis and cDNA microarrays
J. Biol. Chem. 277 (16), 13983-13988 (2002)

JOURNAL

MEDLINE

PUBMED

11825887

Contact: El-Dorri, Hamza

Department of Biochemistry

Institute of Chemistry, University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
BRASIL

Tel: (55) 11-38183848

Fax: (55) 11-38183848

Email: dorri@iq.usp.br

PCR Primers

FORWARD: Universal M13 forward primer

BACKWARD: Universal M13 reverse primer

Plate: 42 row: B column: 10

Seq primer: M13 reverse primer

High quality sequence stop: 409

POLYA=No.

FEATURES

source

1. .719

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM9414 (ATCC26921)"

/db_xref="taxon:51453"

/clone="Tr-A3958"

/sex="Asexual"

/tissue_type="Mycelia"

/lab_stage="18 hr GlyceroI-grown culture"

/db_host="E. coli SOLR cells (kanamycin resistant)"

/clone_lib="TrEST-A"

/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:

XhoI; anamorph: Trichoderma reesei; Cloned

unidirectionally, 5' end of the cDNA cloned into EcoRI

site of pBluescript. Primer: Oligo (dT). Average insert

size: 1.2 kb; Uni-ZAP XR Vector system -5' adaptor

sequence: 5'GAATTCCGACGAG3' -3' adaptor sequence:

5'CTCGAGTTTCTTTTCTTTT3'

ORIGIN

Query Match 11.9%; Score 185; DB 14; Length 719;
Best Local Similarity 67.2%; Pred. No. 1e-32;
Matches 279; Conservative 0; Mismatches 130; Indels 6; Gaps 1;

QY 529 GACCCGTTGTGCGAATGGTGTCTGCGCGGTATCCGGGAGGACTCCCGCGAACCCAG 588
Db 135 GGCCCGGGTTCTCTGATGAGCCTTAACCCGATCCTCCCGACGGGAGCAGCGCCAC 194

QY 589 GTTGTGTGTTCTCAGGGGCGACGGGTATCCTTCGAGTGTTCGGGACGGCAACTCG 648
Db 195 GTCTGTGTTCTCAGGTTCGCGGAGCATCTTGGCCAGGCTCCTTGAGCGGCTACTGCC 254

QY 649 GTCAAAATGGTGTAAAGCGCACCGGCAAGAACACTACTATCCCGGCCAAGATGCCGAT 708
Db 255 CAGGCTGCGGGAACTGGGCGCA-----AAGAACACCGTCACTCCCTGTCAAGGATGCCGAC 308

QY 709 GGAAAGTTCCCTTGGCGGAACTGTAAACAGACTTATCTTCATGCCAAGCATCTCAAGCGC 768
Db 309 GGCAAAATTCCTTGGCGGCTCACTGTCAAAAGAGTATCTGCATGCCAAGCATCTGAAGCGC 368

QY 769 CATCTGTACGCCACACTGGTGTGACCGCGGTACATGTGTCTTTTGAAGACACCTTC 828
Db 369 CACCTGTCTTGGCCACACTGGTGTGATCGCCCTTACATGTGTCTCTGCGGATACCTTC 428

QY 829 TCTGCAATGATATCTGAAACGTCATTTCCAAAATGCTCAATCAGCGGTGTAAACCC 888
Db 429 TCCCGAAGCGACATCTTAAAGCGTCACCTTCAGAAAGGCTCCATTCGCGCGGGAACCC 488

QY 889 ACCGAGCAAGCGACTTGTCCGACCCCAATGGCATGTGAGAGGTGCCAAGC 943
Db 489 ACGGCGGTAGCCATCTCTCACTTAAGCCCCCATGTGAGAGGTGCCAAGC 543

RESULT 3
LOCUS BM077105 776 bp mRNA linear EST 17-NOV-2003
DEFINITION TrEST-A3424 TrEST-A Hypocrea jecorina cDNA clone Tr-A3424 5',
similar to Rfec [Emericella nidulans], mRNA sequence.
ACCESSION BM077105

```

VERSION      BM077105.2 GI:38355949
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Hypocrea jecorina
REFERENCE    Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
AUTHORS      Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,
              Ferreira, J.R.Jr., Aralhao-Neto, J., Farah, J.P.S. and El-Dorry, H.
TITLE        Elucidation of the metabolic fate of glucose in the filamentous
              fungus Trichoderma reesei using expressed sequence tag (EST)
              analysis and cDNA microarrays
JOURNAL      J. Biol. Chem. 277 (16), 13983-13988 (2002)
MEDLINE      21950703
PUBMED       11825887
COMMENT      On Feb 5, 2002 this sequence version replaced gi:18498287.
              Contact: El-Dorry, Hamza
              Department of Biochemistry
              Institute of Chemistry, University of Sao Paulo
              Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
              BRASIL
              Tel: (55) 11-38183848
              Fax: (55) 11-38183848
              Email: dorry@iq.usp.br
              PCR primers
              FORWARD: Universal M13 forward primer
              BACKWARD: Universal M13 reverse primer
              Plate: 36 row: F column: 4
              Seq primer: M13 reverse primer
              High quality sequence stop: 582
              POLYA-No.
              Location/Qualifiers
                1..776
                  /organism="Hypocrea jecorina"
                  /mol_type="mRNA"
                  /strain="QM9414 (ATCC26921)"
                  /db_xref="taxon:51453"
                  /clone="Tr-A3424"
                  /sex="Asexual"
                  /tissue_type="Mycelia"
                  /dev_stage="18 hr Glycerol-grown culture"
                  /lab_host="E. coli SOLR cells (kanamycin resistant)"
                  /clone_lib="rrEST-A"
                  /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                  XhoI; anamorph-trichoderma reesei; Cloned
                  unidirectionally, 5' end of the cDNA cloned into EcoRI
                  site of pBluescript. Primer: Oligo (dT). Average insert
                  size: 1,2 kb; Uni-ZAP XR Vector system -5' adaptor
                  sequence: 5'GAATTCGGCAGG3', -3' adaptor sequence:
                  5'CTCGAGTTTCTTTTCTTTT3'"

FEATURES             source
    source
    1..687
      /organism="Coccidioides immitis"
      /mol_type="mRNA"
      /db_xref="taxon:5501"
      /clone="CIAH91"
      /dev_stage="spherule"
      /lab_host="SOLR"
      /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
      XhoI"

ORIGIN
Query Match      11.0%; Score 171.6; DB 12; Length 776;
Best Local Similarity 60.7%; Pred. No. 1.5e-29;
Matches 318; Conservative 0; Mismatches 199; Indels 7; Gaps 2;

QY 392 TGTCTCCGCGCGCTTCAGCGCTCAGTCGTCACCGGACGCAATTGGCCGCGCGCTCCCC 451
Db 207 TGATGCCCGCCAGGCCACCACTCGCGTCTCTCACCTCAGCGGATTGCTCTCTCGCGCGGTTG 266
QY 452 GGGACCGCTCGTGTGACTTCAACATGCACTTCTCTCAGGAGCATTCAGTTATTCGGACG 511
Db 267 GCGCGCCGGTCTCTCTGTCGTGCTCAATGCTTCCCGTGTGTCATCGCGCAACCCG 326
QY 512 GAAAGCCTCAAGGTGGGACCCCGTGTCTGCGAATGGTGTGCGCCGTATCCCGGGAAGG 571
Db 327 GTGTGGCTCGCCCTACGGCCCGGGTCTCTGTATGAGAGCTTACCCGATCTCTCCGGACG 386
QY 572 ACTCCCCCGGAAACCCAGGTGTGGTGTCTCAGGGGAGCGCGGTATCTTCCGAGTGTTC 631
Db 387 GCGAGCAGCGGACCCACGCTGTGGTGTCTCAGGGTCCCGGAGGATCTTCCCGCGGCTC 446
QY 632 CGGGACGGCGAACTCCGGTCAAAATGGTGTAAACGGCACCGGACGACACTACTATCC 691

```

```

Db 447 CTGGACGGCGCTACT-----GCCAGGCTGGGGAACTGGCGCAAGAACACCGTCATCC 500
QY 692 CGGCCAAGGATCGGATGGAAGTTCCTTGCCTGCGAACTGTAAACAAGACTTATCTTCATG 751
Db 501 CTGTCAAGGATCGGACGCGCAATTCCTTGGCCCTCACTGCACAAGACGATATCTGCATG 560
QY 752 CCAAGCATCTCAAGCGCCATCTGTACGCCACACTGGTGACCGCCCGTACATGTGTGTC 811
Db 561 CCAAGCACCTGAAGCGCCACTGCTTCCGCACACTGGTGATCGCCCTACATGTGTGTC 620
QY 812 TTTCGAAGACACCTTCTCTCGAGTGTATATCTCTGAACGTCATCTTCCAAAATGTCATCA 870
Db 621 TCTGCGCGGATACCTTCTCTCGAAGCGCATCTCTTCCAGGCGTCACTTTCGAAAATGCTTC 680
QY 871 ATCAGGCGTGTAAACCCACCGGAGCAACGCACTTGTTCGACCC 914
Db 681 ATCGCGCGGGCAACCCCAAGGCGCGCTTGGCCCTTTTGGTCCC 724

RESULT 4
BF251517      687 bp      mRNA      linear      EST 15-NOV-2001
LOCUS         EST418778 Coccidioides immitis spherule cDNA library Coccidioides
DEFINITION    immitis cDNA clone CIAH91 5' sequence, mRNA sequence.
ACCESSION     BF251517
VERSION       BF251517.1 GI:16931660
KEYWORDS      EST.
SOURCE        Coccidioides immitis
ORGANISM      Coccidioides immitis
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
              Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE     1 (bases 1 to 687)
AUTHORS       Gardner, M.J. and Kirkland, T.
TITLE         Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL       Unpublished (2000)
COMMENT       Contact: Malcolm J. Gardner
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301 838 3519
              Fax: 301 838 0208
              Email: gardner@tigr.org.
              Location/Qualifiers
                1..687
                  /organism="Coccidioides immitis"
                  /mol_type="mRNA"
                  /db_xref="taxon:5501"
                  /clone="CIAH91"
                  /dev_stage="spherule"
                  /lab_host="SOLR"
                  /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                  XhoI"

ORIGIN
Query Match      7.0%; Score 108.8; DB 10; Length 687;
Best Local Similarity 56.2%; Pred. No. 9.3e-15;
Matches 396; Conservative 0; Mismatches 252; Indels 56; Gaps 8;

QY 95 TGCTCAGTACAGGATGGAGCCTCAGTACCGAGTCTCTCAGCGGACACCGTCCCTCCTC 154
Db 30 TCACCAATATCAGATGTCATTCGCGGTAGCGGATGGCGCAACCGCATAGTTTGCGCCAT 89
QY 155 TGCAACCCCATCATGACCGAGTCGCCGCTCCTCACTCGTACATCGGGGACGCGCGTACC 214
Db 90 TGGACACCATCAGATCCGTGCGCAATTCAGCATACGATATCTC-----CAGCCTTTC 143
QY 215 GGCTGTATCGAACAGGTACCCCGCA---TCAAGTCACGATGTTTACGGGTCTTCTGCTG 271
Db 144 GGAATGACATGCCTAGATACCCGACACACCTCCGGCACCGATGTTTATGGGTATCGTCG 203
QY 272 CGCCGATAATGCCCCACATACGCTGGGAGCTTGCTCCGACATCTTTCCTTTCTCATC 331

```



```

COMMENT      Contact: Dean RA
              Clemson University Genomics Institute
              100 Jordan Hall, Clemson University, Clemson, SC 29634
              Tel: 864 656 5737
              Fax: 864 656 4293
              Email: rdean@clemson.edu
              Seq primer: GGAAACAGCTATGACCATG
              Class: BAC ends
              High quality sequence stop: 236.
              Location/Qualifiers
                1..585
                /organism="Magnaporthe grisea"
                /mol_type="genomic DNA"
                /strain="70-15"
                /db_xref="taxon:148305"
                /clone="mgxb0008J11r"
                /tissue_type="protoplasts"
                /lab_host="E. coli DH10B"
                /clone_lib="CUGI Rice Blast BAC Library"
                /note="Vector: pBACWICH; Site 1: HindIII; Site 2: HindIII;
                Rice blast is one of the most devastating fungal diseases
                of rice world wide. It is a filamentous ascomycete with
                a haploid genome (n=7) of approximately 40 Mbp. Rice
                blast is an important model fungal pathogen for studying
                numerous aspects of the fungal-host interaction. In
                order to facilitate genome wide analysis, a BAC library
                containing 9216 clones with an average insert size of 130
                kbp was constructed. This library represents greater
                than 25X genome coverage. High density colony filters
                are available upon request."

FEATURES             source
  source
    Query Match      5.3%; Score 82; DB 28; Length 585;
    Best Local Similarity 53.2%; Pred. No. 1.8e-08;
    Matches 248; Conservative 0; Mismatches 205; Indels 13; Gaps 3;

Qy 321 CCTTTCTATCCCAATCGGAGGCGAGCGACGCGGAGCAATGCGCGCACTATCTCTCC 380
Db 116 CCCGTACCTGGCATGATGCGCGCGCGAGTACCAACACTCAGCCCGGTATCCAC 175

Qy 381 TCCTCATAGCTGTCTCCGCGCGCTTCAGGCTCAGTGTACCGGAGCAATGGGCC 440
Db 176 CACAACCTCAATGATGACCCAGGATCGACCGCAATGGCCCAACCTCAACCTATTGGCC 235

Qy 441 GCGGCTCCCGGAGCGCTGCTGACTTCAACAATGGACTTCTTCAGGAGCATTCAG 500
Db 236 CGCGCTCCTGTGTGCGCTCCCGGTTATGGTCCAAATGCCGGGTGGCATGGT 295

Qy 501 TTAATCGGACGGAAG-----CCTCAAGTTGGACCCCGTTGCTGGAATGTTGCTGCG 555
Db 296 CCCTCAGGGCGGCATGTGTCGCATACGCGCAGAGCCCATGATGCCAG--CCTCAA 353

Qy 556 CCGTATCCGGAGAGTATCCCGGACCCGACCCAGTGTGTGTTCTCAGGGCGAGCGGT 615
Db 354 ATGATCAACAGGAAGCGACGCGCGGACACATGTGTTGGCTCTCAGGGTTCGCGGAG 413

Qy 616 ATCTCTCCGAGTGTTCGGGAGCGCAACTCCGCTCACAATGTTGTTAAGCGCACCGGC 675
Db 414 ATCTCTCCGAGCGCGCGGTAGG-----CCGAGGCCCGGTGCTGGCAGCCAGCG 467

Qy 676 AAGAACTACTATCCCGGCGCAAGGATCCGATGGAAGTTCCCTTGCCCGAATGTAAC 735
Db 468 GCAAGAACAACATATCCCGGACAGGATGCGGAGGTTCCCGTGCCTCATTTGTAAC 527

Qy 736 AAGACTTATCTTATGCAACAGTCTCAGCGCATCTGTACGCC 781
Db 528 AAGACGTACCTGCAAGCAAGCATCTCAAGAGGATCTTCTGAGAC 573

ORIGIN
  Query Match      3.6%; Score 56.2; DB 13; Length 807;
  Best Local Similarity 31.5%; Pred. No. 0.026;
  Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;

Qy 15 CCATCCCTCTCGGCTCGCTTACAGCTTGCCTCAAGGATCGGGGCTCTTCTCTTGC 74
Db 230 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 289

Qy 75 TCCCATCTCGAGCATGCGCAATGCTCAGTACACATGACGCTCAGTACCGAGTCTCTCA 134
Db 290 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 349

Qy 135 GCGGCACACCTGTCTCTGTGCAACCCCATATAGCAGTGTGCGGCTCTCTACTGTA 194
Db 350 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 409

Qy 195 CATGGGACGCGCGTACCGGCTGATGTGAACAGGTACCCGCTACAGTACAGTGT 254
Db 410 SCGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 469

Qy 255 TTACGGCTCTTCTGTGCGCGGATATATGCCCCACATACCGTGGGAGCTTGCCTCGAC 314
Db 470 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 529

Qy 315 ATCTTCTCTTCTCATGCCAATCGGAGGCGGACAGGCGAGCAATGCGGCACTA 374
Db 530 CCGCGCCSCCGCCSCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 589

Qy 375 TCCTCTCTCTCATAGGTGCTCCGCGCGCTTCCAGGCTCAGTGTACCGGAGCAAT 434
Db 590 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 649

Qy 435 TGCAGCGCGCTCCCGGAGCGGTGCTGCTGACTTCAACAATGGACTTCTCTTCAAGAGC 494

```


715	TTCCGCTGCCGGAACCTGTATACAAGACTTATCTTCATGCGAAGCATCTCAAGCGGCACTCG	774
QY		
471	TACACATGCCGCCGAGTGTACGGCTCTTCAACACAGCGAGCCATCTCAAGTCCCAATG	530
Db		
775	CTACGGCACACTGGTGTACCGCCGGTACANGTGTCTTTTGCAAGACACCTTCTCTGCG	834
QY		
531	CGCCTCCACACGGGTGAGCGGCGGTACAAGTGCCAGCAATGTGCAAGTCCCTTCAACACC	590
Db		
835	AGTGATATCTGAAAAAGTCATTTTCCAAAAATGCTCAATCAGGGCGTGTAAACCCCAACC	891
QY		
591	AACGCTGAGTCTGAAAAGCCACGTCACACGGCTACACCCAGGCGCTTGGCTACGGAGCC	647
Db		

RESULT 11	CG748837/c	CG748837	1970 bp	DNA	linear	GSS 24-OCT-2003
LOCUS						
DEFINITION						

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE	Neodiprogasteridae; Pristionchus. 1 (bases 1 to 1970)
AUTHORS	Srinivasan, J., Sinz, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K., Buntjer, J., van der Meulen, M. and Sommer, R. J.
TITLE	An integrated physical and genetic map of the nematode <i>Pristionchus</i> <i>pacificus</i>
JOURNAL	Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE	22835951
PUBMED	12884007

Accepted, 12/1/00
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

```

FEATURES
    source
        Tel: 00497071601371
        Fax: 00497071601498
        Email: ralf.sommer@tuebingen.mpg.de
        Class: BAC ends.
        Location/Qualifiers
            1..1970
                /organism="Pristionchus pacificus"
                /mol_type="genomic DNA"
                /strain="California"
                /db_xref="taxon:54126"
                /clone_lib="ppa EcoRI BAC Library"
                /note="The library was generated by a partial digest of
                        the genomic DNA with EcoRI and cloning into the BAC
                        vector"

```

```

/clone.lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector"

```

db	895	CNCCCCCCCNCNCNCCCCCNCCNNNNNCCNNCCNCCCCCCCCCNC	836
Qy	195	CATGGGCGCGCGTACCGCCTGTACTGAACAGGTACCGGCATCAAGTCAGATGT	254
db	835	CNNCNCNNCCCCCNCCNCCNCCNCCNCCCCCCCCCNCNCCNCCNCCNCCNCC	776

[illegible]

[illegible]

```
QY 79 ATCTGAGCATGGCAATGCTCTAGTACAGATGAGGCTCAGTACCCAGTCTCTCAGCGG 138
Db 670 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 611
QY 139 CACACCTGTGCTCTCTGCAACCCCATATAGCCAGTGGCGCTCCTCACTGCTATG 198
Db 610 CCCCCCNNGCCNCCCCCCCCCCCCCCCCCNCCNCCNCCNCCNCCNCCNCCNCCN 551
QY 199 GGGCAGCGCGGTACCGGCTGTATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTAC 258
Db 550 CCNCCGCCGCCGCCNCCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 491
QY 259 GCGTCTTCTGTGCGCGGATATGCCCCACACTACCGTGGCAGCTTGCTCCGACATCT 318
Db 490 CCCCCNCCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 431
QY 319 TTCCTTTCTCATGCCAATCCGCGAGGCGCAGGCACAGGCGAGCAATCGCGGCACTATCCT 378
Db 430 CCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 371
QY 379 CCTCTCATAGCGGTGCTCCGCGTTCCAGCGCTCAGTCTGTAACCGCAGCCAAATTGG 438
Db 370 CCCCCNCCNCCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 311
QY 439 CCGGCGCTCCGC 451
Db 310 CCCCCCCCCNCC 298
```

Search completed: April 3, 2004, 11:01:06
Job time : 3732 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:38:48 ; Search time 59 Seconds
(without alignments)
2485.460 Million cell updates/sec

Title: US-10-029-180-8

Perfect score: 2866

Sequence: 1 MDPNHSRPSTSLPQSSA.....NGLYLASTMSDGTGVQPARQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1596107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1596107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2866	100.0	519	5	ABP63087 FLO11 gen
2	2866	100.0	600	5	ABP63143 Fungal ge
3	2866	100.0	611	5	ABP63142 Fungal ge
4	285.5	8.9	722	4	AAW39029 Human pol
5	251.5	8.8	2703	4	ABP63299 Drosophil
6	251.5	8.8	2703	5	ABG70019 Larval vi
7	249	8.7	554	4	ABG15134 Novel hum
8	240	8.4	723	7	ABG64092 Human pro
9	240	8.4	2062	7	ADC31187 Human nov
10	239	8.3	1401	4	ABG24037 Novel hum
11	238.5	8.3	1142	7	ADC07968 Rice prot
12	237	8.3	1192	4	ABP65293 Drosophil
13	236.5	8.3	644	4	ABP70436 Drosophil
14	233.5	8.1	1185	2	AAW33497 Human atr
15	229.5	8.0	926	4	ABP65135 Drosophil
16	223	7.8	420	5	ABP57247 Mouse isc
17	211	7.4	1963	4	ABP62819 Drosophil
18	210.5	7.3	505	4	AAW61923 Human ear
19	210.5	7.3	543	2	AAW24022 Human pro
20	210.5	7.3	543	2	AAW76985 Human EGR
21	210.5	7.3	543	3	AAW51116 Human EGR
22	210.5	7.3	543	4	AAU09066 Human ZIF
23	210.5	7.3	543	4	AAW76861 Human lun
24	210.5	7.3	543	5	AAU85516 Clone #18
25	210.5	7.3	543	6	ABU69488 Human lun

26	210.5	7.3	543	6	ABU66390 Lung canc
27	210.5	7.3	543	7	ABP75271 Prostate
28	210.5	7.3	543	7	ADC37141 543 amino
29	210.5	7.3	543	7	ADD48114 Human Pro
30	210.5	7.3	543	8	ADE71608 Human lun
31	209.5	7.3	496	2	AAW76984 Mouse Egr
32	209.5	7.3	533	1	AAW93113 Egr-1.3/
33	209.5	7.3	533	2	AAW63129 Mouse Egr
34	209.5	7.3	533	4	AAW78054 Human zin
35	209.5	7.3	533	5	ABP57356 Mouse isc
36	209	7.3	398	5	AAW49116 Mouse Sp1
37	207.5	7.2	398	5	AAW49117 Human Sp1
38	207.5	7.2	398	7	ADC31406 Human nov
39	207.5	7.2	978	7	ADC31742 Human nov
40	207.5	7.2	1386	7	ADE59291 Rat Prote
41	207.5	7.2	1386	7	ADE59295 Rat Prote
42	207	7.2	488	5	AAO20509 Protein o
43	206	7.2	428	5	AAE25076 Mouse ost
44	206	7.2	551	7	ADC31398 Human nov
45	205.5	7.2	345	2	AAW52577 Human tra

ALIGNMENTS

RESULT 1
ABP63087

ID ABP63087 standard; protein; 519 AA.

XX AC ABP63087;

XX XX

DT 22-OCT-2002 (first entry)

XX XX

DE FLO11 gene expression regulator An09.

XX XX

Fungal gene expression regulator; fungicide; gene therapy; An01; An05; An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05; At07; At08; At11; At14; At16; At18; At19; At22; At24; At27; At32; PC05; PC06; PC07; PC08; PC09; PC10; PC18; PC23; PC24; PC25; PC33; PC34; FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320; At322; An100; At167; At221; At233; At240; At274; PC1000; PC1001; lovF; lovE; lovastatin; PC804; acvA; penicillin; antifungal.

XX OS Unidentified.

XX XX

FN W0200257456-A2.

XX XX

PD 25-JUL-2002.

XX XX

PF 24-DEC-2001; 2001WO-US049911.

XX XX

PR 22-DEC-2000; 2000US-0257431F.

XX XX

PA (MICR-) MICROBIA.

XX XX

PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

PI Holtzman D, Sherman A;

XX XX

DR WPI; 2002-627365/67.

XX XX

DR N-PSDS; ABQ94220.

XX XX

PT New isolated or recombinant gene, or purified protein, useful in

PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the

XX XX

PS medically useful products.

XX XX

PS Claim 57; Page 37; 71pp; English.

XX XX

CC The present invention relates to novel fungal gene expression regulators

CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,

XX XX

CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,

CC At14, At16, At18, At19, At20, At22, At24, At27, At32, PC05, PC06, PC07,

CC PC08, PC09, PC10, PC18, PC23, PC24, PC25, PC33 and PC34 are FLO11 gene

CC expression regulators. FLO11 is required for fungal invasion and its
 CC expression is believed to be regulated by factors that also modulate
 CC secondary metabolite production. At279, At286, At291, At320, At322,
 CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
 CC LOVF gene expression regulators, and At501 and At574 are LOVF gene
 CC expression regulators. lovF and lovE are believed to be involved in the
 CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
 CC expression regulator. acvA is involved in the production of the secondary
 CC metabolite penicillin. The fungal gene expression regulators and their
 CC coding sequences are useful in regulating or manipulating the expression
 CC of fungal genes that are involved in the production of enzymes, secondary
 CC metabolites and other commercially and medically useful products, in
 CC order to achieve maximum benefit. The genes may also be used to identify
 CC genes relevant to fungal invasion which may act as targets for the
 CC development of antifungal drugs
 CC
 CC SQ Sequence 519 AA;

Query Match 100.0%; Score 2866; DB 5; Length 519;
 Best Local Similarity 100.0%; Pred. No. 5.9e-193;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRNHPSPSTSLPQGSAPLPSAPISMPMPQYTMQPPVPSQPHTLPLQPHHSQSP 60
 DB 1 MDRNHPSPSTSLPQGSAPLPSAPISMPMPQYTMQPPVPSQPHTLPLQPHHSQSP 60

QY 61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPTMPHTTVGSLPPTFLSHPNPQAAQ 120
 DB 61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPTMPHTTVGSLPPTFLSHPNPQAAQ 120

QY 121 QSPHYPPHSLVLPASSAQSYPOPIAPAPRRDRADFNNGLPAGFYSYDGKPGQWDVPA 180
 DB 121 QSPHYPPHSLVLPASSAQSYPOPIAPAPRRDRADFNNGLPAGFYSYDGKPGQWDVPA 180

QY 181 ANGAAPYPGKDSPTQVGSQGRGILPSVPGRATPVNGTGNKNTTIPAKDAGKFP 240
 DB 181 ANGAAPYPGKDSPTQVGSQGRGILPSVPGRATPVNGTGNKNTTIPAKDAGKFP 240

QY 241 CPNCKNTYLHAKHLRHLLRHTGDRPVMCVLCKDTFSRSDILKRHFQKCSIRGNTGAT 300
 DB 241 CPNCKNTYLHAKHLRHLLRHTGDRPVMCVLCKDTFSRSDILKRHFQKCSIRGNTGAT 300

QY 301 HLSHPNAHVKSQQQAAANPVKPVQDEVSTVPPNGIPCTTYGEGAVNGNGLAPAPGY 360
 DB 301 HLSHPNAHVKSQQQAAANPVKPVQDEVSTVPPNGIPCTTYGEGAVNGNGLAPAPGY 360

QY 361 ADHQTWGFPMSSVNGMGRGQPEDAFFGGRPHQAPWPQKQSPYLVPQADPSGHQINI 420
 DB 361 ADHQTWGFPMSSVNGMGRGQPEDAFFGGRPHQAPWPQKQSPYLVPQADPSGHQINI 420

QY 421 DRNIEVKQPVVODPKRPVMPGHPGHPGLDWTSMFQPAPEGYSQSMGQOEPIHAH 480
 DB 421 DRNIEVKQPVVODPKRPVMPGHPGHPGLDWTSMFQPAPEGYSQSMGQOEPIHAH 480

QY 481 VETERXYTTTAGQSGMNGLYLASTMSGDGTVPAPQ 519
 DB 481 VETERXYTTTAGQSGMNGLYLASTMSGDGTVPAPQ 519

RESULT 2
 ID ABP63143
 XX ABP63143 standard; protein; 600 AA.

AC ABP63143;

XX 22-OCT-2002 (first entry)

XX Fungal gene expression regulator SEQ ID 120.

DE Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
 KW An07; An10; An13; An17; An28; An34; At01-1; At01-2; At03; At05;
 KW At09; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
 KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;

KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
 KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
 XX lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal.
 OS Unidentified.
 XX WO200257456-A2.
 XX 25-JUL-2002.
 XX 24-DEC-2001; 2001WO-US049911.
 XX 22-DEC-2000; 2000US-0257431P.
 XX (MICR-) MICROBIA.
 XX
 XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
 PI Holtzman D, Sherman A;
 PI WPI; 2002-627368/67.
 XX
 XX New isolated or recombinant gene, or purified protein, useful in
 PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
 PT production of enzymes, secondary metabolites or other commercially and
 PT medically useful products.

XX Claim 67; Page 67; 71pp; English.

CC The present invention relates to novel fungal gene expression regulators
 CC (AB094217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
 CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
 CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
 CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
 CC expression regulators. FLO11 is required for fungal invasion and its
 CC expression is believed to be regulated by factors that also modulate
 CC secondary metabolite production. At279, At286, At291, At320, At322,
 CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
 CC LOVF gene expression regulators, and At501 and At574 are LOVF gene
 CC expression regulators. lovF and lovE are believed to be involved in the
 CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
 CC expression regulator. acvA is involved in the production of the secondary
 CC metabolite penicillin. The fungal gene expression regulators and their
 CC coding sequences are useful in regulating or manipulating the expression
 CC of fungal genes that are involved in the production of enzymes, secondary
 CC metabolites and other commercially and medically useful products, in
 CC order to achieve maximum benefit. The genes may also be used to identify
 CC genes relevant to fungal invasion which may act as targets for the
 CC development of antifungal drugs

XX Sequence 600 AA;

Query Match 100.0%; Score 2866; DB 5; Length 600;
 Best Local Similarity 100.0%; Pred. No. 7e-193;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRNHPSPSTSLPQGSAPLPSAPISMPMPQYTMQPPVPSQPHTLPLQPHHSQSP 60
 DB 82 MDRNHPSPSTSLPQGSAPLPSAPISMPMPQYTMQPPVPSQPHTLPLQPHHSQSP 141

QY 61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPTMPHTTVGSLPPTFLSHPNPQAAQ 120
 DB 142 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPTMPHTTVGSLPPTFLSHPNPQAAQ 201

QY 121 QSPHYPPHSLVLPASSAQSYPOPIAPAPRRDRADFNNGLPAGFYSYDGKPGQWDVPA 180
 DB 202 QSPHYPPHSLVLPASSAQSYPOPIAPAPRRDRADFNNGLPAGFYSYDGKPGQWDVPA 261

QY 181 ANGAAPYPGKDSPTQVGSQGRGILPSVPGRATPVNGTGNKNTTIPAKDAGKFP 240
 DB 262 ANGAAPYPGKDSPTQVGSQGRGILPSVPGRATPVNGTGNKNTTIPAKDAGKFP 321

QY 241 CPNCKNTYLHAKHLRHLLRHTGDRPVMCVLCKDTFSRSDILKRHFQKCSIRGNTGAT 300
 DB 321 CPNCKNTYLHAKHLRHLLRHTGDRPVMCVLCKDTFSRSDILKRHFQKCSIRGNTGAT 300

Db 322 CPNCNTYILHAKHLKHLLEHTGDRPYMCVLCNDTFSRSDILKRHPQKCSIRRGNTGAT 381
 QY 301 HLSHPNAHVKRSQQQAANPVKPVQDEVSTVPPNGIPGTTTGEVANGVGLAPARPGY 360
 Db 382 HLSHPNAHVKRSQQQAANPVKPVQDEVSTVPPNGIPGTTTGEVANGVGLAPARPGY 441
 QY 361 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQAGAPWQAPKQSPYLQVQFADPSGHQLNI 420
 Db 442 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQAGAPWQAPKQSPYLQVQFADPSGHQLNI 501
 QY 421 DRNIEQVKQPVQDPKRPVMPGHPGHELDWTSMFOQAPEGYMFSQMPGQGPPIHAH 480
 Db 502 DRNIEQVKQPVQDPKRPVMPGHPGHELDWTSMFOQAPEGYMFSQMPGQGPPIHAH 561
 QY 481 VETERKYPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 519
 Db 562 VETERKYPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 600
 RESULT 3
 ID ABP63142
 XX ABP63142 standard; protein; 611 AA.
 AC ABP63142;
 XX
 DT 22-OCT-2002 (first entry)
 XX
 DE Fungal gene expression regulator SEQ ID 119.
 XX
 KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
 An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
 At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
 Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
 FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
 At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
 lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal.
 XX
 OS Unidentified.
 XX
 PN WO200257456-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 24-DEC-2001; 2001WO-US049911.
 XX
 PR 22-DEC-2000; 2000US-0257431P.
 XX
 PA (MICR-) MICROBIA.
 XX
 PI Cali BW, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
 PI Holtzman D, Sherman A;
 XX
 DR WPI; 2002-627368/67.
 XX
 PT New isolated or recombinant gene, or purified protein, useful in
 PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
 PT production of enzymes, secondary metabolites or other commercially and
 PT medically useful products.
 XX
 PS Claim 67; Page 67; 71pp; English.
 XX
 CC The present invention relates to novel fungal gene expression regulators
 CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
 CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
 CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
 CC Pc08, Pc09, Pc10, Pc18, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
 CC expression regulators. FLO11 is required for fungal invasion and its
 CC expression is believed to be regulated by factors that also modulate
 CC secondary metabolite production. At279, At286, At291, At320, At322,
 CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
 CC lovF gene expression regulators, and At501 and At574 are lovE gene
 CC expression regulators. lovF and lovE are believed to be involved in the
 CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene

CC expression regulator. acvA is involved in the production of the secondary
 CC metabolite penicillin. The fungal gene expression regulators and their
 CC coding sequences are useful in regulating or manipulating the expression
 CC of fungal genes that are involved in the production of enzymes, secondary
 CC metabolites and other commercially and medically useful products, in
 CC order to achieve maximum benefit. The genes may also be used to identify
 CC genes relevant to fungal invasion which may act as targets for the
 CC development of antifungal drugs
 XX
 SQ Sequence 611 AA;

Query Match 100.0%; Score 2866; DB 5; Length 611;
 Best Local Similarity 100.0%; Pred. No. 7.1e-193;
 Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPNRHPSRPSTSLPQGSAPLPSAPISSMMPMPQYTMQDPYVQSHLTPPLQPHHSQSP 60
 Db 93 MDPNRHPSRPSTSLPQGSAPLPSAPISSMMPMPQYTMQDPYVQSHLTPPLQPHHSQSP 152
 QY 61 APHSYMGOPPYVRPDLNRYPASSHDVYASSAAPIMPHTTVGSLLPPTSFSLHPNPQAAQAO 120
 Db 153 APHSYMGOPPYVRPDLNRYPASSHDVYASSAAPIMPHTTVGSLLPPTSFSLHPNPQAAQAO 212
 QY 121 QSPHYPPPHSVLPPASSAQSYEQPIAPAPRRDRADFNNGLPFGAFSYSDGKPGQWDVPA 180
 Db 213 QSPHYPPPHSVLPPASSAQSYEQPIAPAPRRDRADFNNGLPFGAFSYSDGKPGQWDVPA 272
 QY 181 ANGAAPYFKGDSPTQTVVGSQGRGILPSVPGRATPVTVNGVNGTKNTTIPAKDADGKPP 240
 Db 273 ANGAAPYFKGDSPTQTVVGSQGRGILPSVPGRATPVTVNGVNGTKNTTIPAKDADGKPP 332
 QY 241 CPNCNTYILHAKHLKRHLRHTGDRPYMCVLCNDTFSRSDILKRHPQKCSIRRGNTGAT 300
 Db 333 CPNCNTYILHAKHLKRHLRHTGDRPYMCVLCNDTFSRSDILKRHPQKCSIRRGNTGAT 392
 QY 301 HLSHPNAHVKRSQQQAANPVKPVQDEVSTVPPNGIPGTTTGEVANGVGLAPARPGY 360
 Db 393 HLSHPNAHVKRSQQQAANPVKPVQDEVSTVPPNGIPGTTTGEVANGVGLAPARPGY 452
 QY 361 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQAGAPWQAPKQSPYLQVQFADPSGHQLNI 420
 Db 453 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQAGAPWQAPKQSPYLQVQFADPSGHQLNI 512
 QY 421 DRNIEQVKQPVQDPKRPVMPGHPGHELDWTSMFOQAPEGYMFSQMPGQGPPIHAH 480
 Db 513 DRNIEQVKQPVQDPKRPVMPGHPGHELDWTSMFOQAPEGYMFSQMPGQGPPIHAH 572
 QY 481 VETERKYPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 519
 Db 573 VETERKYPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 611

RESULT 4
 AAM39029

ID AAM39029 standard; protein; 722 AA.

XX
 AC AAM39029;

XX
 DT 22-OCT-2001 (first entry)

XX
 DE Human polypeptide SEQ ID NO 2174.

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Sisydrager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 XX leukaemia.

OS Homo sapiens.

XX
 PN WO200153312-A1.

XX

Db 292 PQLGFSQQYRTPPTNTSRGQSPYPAHQNSGVSFPSSPQQQQQQQQQQQQQQQQQQGCP 351
 Qy 46 -PHILPP---LQPHSQSPAHSYMGQPPYRDLNRYPA--SSH-DVYASSAAPIMPHTT 98
 Db 352 VFGGPPGTGQQPPQOQNTPTTSQSPYQRYPTTPEGLPAGGSHRTAYSTHQYEPNRPW 411
 Qy 99 VGSLLPPTSLFSLHNPQQAQAQQSPHY-----PPHLSVL--PPASSAQSY-----P 142
 Db 412 PGGSSPSPGSHPLDPA-----SPHHVPTLQQPPPPHVSAGPPSSSPGSHAPSP 465
 Qy 143 QP--IAPAPRRDRADFNGLPSGAFSYDGKPGQWDEVAANGAAPYPGKDSPTQV-- 198
 Db 466 QPSQASPSPHQLIQNSDSSG-----GAHSGM-----GSGP-PGTENPQ-QVWRP 511
 Qy 199 -----GSGRGLILPSYPCR--ATPVTVNGVNGTKNTTIPAKDADGKFCPCNCKTYLH 250
 Db 512 TSPSTGSSGSRSMSPAVAQNHPISRPASN-QSSSGGPMQPPVAGGPPMP----- 562
 Qy 251 AKHLKRLRLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKSIIRGNPTGATHLSHPNAHV 310
 Db 563 -----PH-----PGMPGPP-----PQQQ 575
 Qy 311 RSQQAAN-----PVKPVQDEVSTVPPNGIPGTTY-----GEG 346
 Db 576 QSQQQASNSASSASNSPQOTPPAPPNQCMMNMTATPPPPQAGGYPMPHMHGGY 635
 Qy 347 AVNGGLAPARPVADHQTMGF-----GHPGELDWTSMFOQAPEGY-----MFSQ 468
 Db 636 KMGGPGSGAGGYPQPPQYPPGNYPPRPPQYPPGAYATGPPPTTSQAGAGGANSMP 695
 Qy 383 DAFPGGRPHQCAP-----W-POAPKQSPYLVQPGADPSGHQLNIDRNIQV-- 428
 Db 696 GAQAGGYPGRGMNHTGQYPPYQWVPPSPQOT-----VPGGAPGAMVG---NHVQKGT 748
 Qy 429 -QPVVQDPKRPVMPGHP-----GHPGELDWTSMFOQAPEGY-----MFSQ 468
 Db 749 PPPVVGPPPPQSGSRPRLNYLKQHLQHKGGYGG-----SPTPQGGYNGGTGMHP 803
 Qy 469 SMPGQGPPIH---AHVETERKYPTTTAGQESGMNG 501
 Db 804 GMPMG-PPHMGPPHGTNMG-PPTSTPPQSQMLQG 837

RESULT 6

ABG70019
 ID ABG70019 standard; protein; 2703 AA.
 AC ABG70019;

XX 05-NOV-2002 (first entry)
 DT

XX Larval viability associated protein #18.
 DE

XX Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
 KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;
 KW oilseed rape; soybean; vegetable crop; fruit.
 XX

OS Drosophila melanogaster.
 XX

XX WO200257455-A2.
 PN

XX 25-JUL-2002.
 PD

XX 18-JAN-2002; 2002WO-US001568.
 PF

XX 18-JAN-2001; 2001US-0262351P.
 PR

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA

XX Stam L, Bachmann J, Broadus J, Kamdar KP;
 FI

XX WPI; 2002-590746/63.
 DR

DR N-PSDB; ABS51395.
 XX

PT Identifying inhibitors of activity of proteins essential for *Drosophila*
 larval viability comprises expressing in a host a protein essential for
 larval activity and identifying compounds that inhibit or interact with
 the protein.
 PT

XX Claim 1; Page 108-117; 169pp; English.
 PS

CC The invention describes a method of identifying compounds that inhibit
 the activity of, or that interact with a protein essential for *Drosophila*
 larval viability comprising expressing in a recombinant host a DNA
 molecule to produce a protein essential for larval viability. The method
 is useful for identifying compounds with insecticidal activity. Compounds
 identified are useful as insecticides in crops such as maize, wheat,
 oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar
 beet, oilseed rape, soybeans, vegetable crops and fruits. This is the
 amino acid sequence of a fruit fly larval viability associated protein
 CC

XX Sequence 2703 AA;
 SQ

Query Match 8.8%; Score 251.5; DB 5; Length 2703;
 Best Local Similarity 23.9%; Pred. No. 1.5e-08;
 Matches 152; Conservative 49; Mismatches 208; Indels 227; Gaps 36;

Qy 3 PRNIPSR-----PPSTSLPQGSAPLPSA-----PISSMPMPQYTMQPVVVSQ----- 45

Db 292 POLGFSQQYRTPPTNTSRGQSPYPAHQNSGVSFPSSPQQQQQQQQQQQQQQQQQQGCP 351

Qy 46 -PHILPP---LQPHSQSPAHSYMGQPPYRDLNRYPA--SSH-DVYASSAAPIMPHTT 98

Db 352 VFGGPPGTGQQPPQOQNTPTTSQSPYQRYPTTPEGLPAGGSHRTAYSTHQYEPNRPW 411

Qy 99 VGSLLPPTSLFSLHNPQQAQAQQSPHY-----PPHLSVL--PPASSAQSY-----P 142

Db 412 PGGSSPSPGSHPLDPA-----SPHHVPTLQQPPPPHVSAGPPSSSPGSHAPSP 465

Qy 143 QP--IAPAPRRDRADFNGLPSGAFSYDGKPGQWDEVAANGAAPYPGKDSPTQV-- 198

Db 466 QPSQASPSPHQLIQNSDSSG-----GAHSGM-----GSGP-PGTENPQ-QVWRP 511

Qy 199 -----GSGRGLILPSYPCR--ATPVTVNGVNGTKNTTIPAKDADGKFCPCNCKTYLH 250

Db 512 TSPSTGSSGSRSMSPAVAQNHPISRPASN-QSSSGGPMQPPVAGGPPMP----- 562

Qy 251 AKHLKRLRLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKSIIRGNPTGATHLSHPNAHV 310

Db 563 -----PH-----PGMPGPP-----PQQQ 575

Qy 311 RSQQAAN-----PVKPVQDEVSTVPPNGIPGTTY-----GEG 346

Db 576 QSQQQASNSASSASNSPQOTPPAPPNQCMMNMTATPPPPQAGGYPMPHMHGGY 635

Qy 347 AVNGGLAPARPVADHQTMGF-----GHPGELDWTSMFOQAPEGY-----MFSQ 382

Db 636 KMGGPGSGAGGYPQPPQYPPGNYPPRPPQYPPGAYATGPPPTTSQAGAGGANSMP 695

Qy 383 DAFPGGRPHQCAP-----W-POAPKQSPYLVQPGADPSGHQLNIDRNIQV-- 428

Db 696 GAQAGGYPGRGMNHTGQYPPYQWVPPSPQOT-----VPGGAPGAMVG---NHVQKGT 748

Qy 429 -QPVVQDPKRPVMPGHP-----GHPGELDWTSMFOQAPEGY-----MFSQ 468

Db 749 PPPVVGPPPPQSGSRPRLNYLKQHLQHKGGYGG-----SPTPQGGYNGGTGMHP 803

Qy 469 SMPGQGPPIH---AHVETERKYPTTTAGQESGMNG 501

Db 304 GMPMG-PPHMGPPHGTNMG-PPTSTPPQSQMLQG 837

RESULT 7

ABG15134
 ID ABG15134 standard; protein; 554 AA.


```

XX AC ABG15134;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #15125.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS79321.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 45493; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 554 AA;

Query Match      8.7%; Score 249; DB 4; Length 554;
Best Local Similarity 21.6%; Pred. No. 3.7e-09;
Matches 125; Conservative 44; Mismatches 155; Indels 254; Gaps 27;

QY 10 PPSTS-----LPQGSAPLP--SAPI-SSMPMPQVTMQP-----QYVPSQPHITLPPQ 53
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 PPAASREVWLAGSCPLPLESSPAHRAVERPSVAHQPAVPCARRSLEVPSVEPLTVPALT 70
QY 54 PHSQSPAPHSYNGQPPYREDLNRYPASSHDVYASSAAPIMPTTGVSLPPTSFSLGHPNP 113
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 P-----PRIPAMLTAVCGSL-----PRDRADFNG- 160
QY 114 QAAQAQQSQSPHYPPHSLVPPASSAQ-----STPQPTAPAP-----PRDRADFNG- 160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 86 --GSQHTAPHASPPRLDLQLQTYQCHTSPKAGDYSPLQPGRLQLPLGPEYDFSQGY 143
QY 161 -----LPSGAFS-----YSDGKPGW---DPVAANGA--A 185
DB 144 ELFGASSRVTCEDLESPLAPGPFSSKLQPDMSHY-----ESWFFPTHGADGSGWMD 198
QY 186 PYFG--KDSPTQ--VVGSGQRRG-----ILPSVPCR-- 213
DB 199 LHEGTSWMDLPHTOGALTSPCHPGALQAGLGVYGDHQLCAPPHPHAHLLPAAGQHL 258
QY 214 -----ATPVTVNGVT-----GKNTTIPAKDADGKPPCPN----- 243
DB 259 LGPPDGAKALEVAAPESQGLDSDLDGAARPKGRRSVPRSSGQTVCRPCNCLAEERLGP 318
QY 244 -----CNKTYLHAKHLKRLHRTGDRPYMC--VLCKDTSRSDILX 283
DB 319 CGPDGGKKHLHNCHPGCGKAYAKTSHLKAHLRWHSGDRPFVNCNMLFCGKRTRSDQLQ 378
QY 284 RHPG-----KCSI-----RRGNPTGATHLSHP 305
DB 379 RHLQTHGTGKFFCAVCSRVFMSDLAKHMKTHEGAKEEAAGASGEGKAGGA--VEPP 436
QY 306 NAHVXRSQ--QAAANPVKPVQDEVSTVPPNGIP-----GTTYGEGAVNGNLAP 355
DB 437 GKGKREAEAGSMASSPDSCDCFCVSVPPASAIPAVIFAHELGPTPGRGA--GGGVCP 494
QY 356 ARPGYADHQTMTGPFMSVNGMGEGOPED-----AFPGGR 389
DB 495 RRHSHWEPAPGLXAVGGGGGASGQAGDRGCHAXEAGR 532

RESULT 8
ADB64092
ID ADB64092 standard; protein; 723 AA.
XX AC ADB64092;
XX DT 04-DEC-2003 (first entry)
XX DE Human protein encoded by clone BRAWH20176950.
XX KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
XX KW cell regeneration; membrane protein; signal transduction-related protein;
XX KW transcription-related protein; osteoporosis; neurological disease;
XX KW cancer; tumour.
XX OS Homo sapiens.
XX PN EP1308459-A2.
XX PD 07-MAY-2003.
XX PF 28-MAR-2002; 2002EP-00007401.
XX PR 05-NOV-2001; 2001JP-00379298.
XX PR 25-JAN-2002; 2002US-00350978.
XX PA (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX DR WPI; 2003-450961/43.
XX DR N-PSDB; ADB62122.
XX PT New polynucleotides and polypeptides, useful for developing a diagnostic
XX PT marker or medicines for regulation of their expression and activity, or
XX PT as targets of gene therapy.
XX PS Claim 1; Page; 222pp; English.
XX

```

CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

XX Sequence 723 AA;

Query Match 8.4%; Score 240; DB 7; Length 723;
Best Local Similarity 22.6%; Fred. No. 2.1e-08;
Matches 132; Conservative 54; Mismatches 198; Indels 200; Gaps 27;
QY 19 SAPLPSAPISAMPPOYTMQPOYVPSQPHHTLPPLQPHHSQSPHVSVMGQPPYRDLNRY 78
Db 84 SMPFPFVPSQMPPLPMPP--FVLPSLSPPVMP--PALPATVPPGMP-----131
QY 79 PASSHDVYASSAAMPHTTGVSLPPTSLFSLHNPQAQAQQSPHVPPPH-----SV 131
Db 132 -----FVMPPLSLTSPVPPGM-----PPSLSSAGPPVLPPLPSLSAGPPV 173
QY 132 LPPASSAQSYQPTAPAPRRERADFNGLPSGAFSYDGPQWDVPAANGAPYPGKD 191
Db 174 LPPPLSLSTAPPVPMPLPLPLS-SATPPPGIP-----PPGVPPQGTTPQLT--AAPVPPAS 224
QY 192 SPRTQVVSQGRGILPSVGRATPVINGVNGTKNTTI-----PAKDAGKFPQCN 243
Db 225 SSQSQWPEKEPALLP-----TPVS--FGSAPPTTHPPLQSAGSEQVNSKAP---272
QY 244 CNKTYLHAKHLKRHLRHTGDRPYMCVLCKDTFSRDLKRHFQKCSIRRGNTGATHLS 303
Db 273 -----LSKSAJ-----PY-----SSFSSDQGL-----289
QY 304 HPNAHVKSQQQAANPKPVQD-----EVSSTVPPNGIPGTYGEG-----346
Db 290 -----GESSAAPSQPTAVKDMFVRSGLLPDP-----PRSSYLESPRGPRFED 339
QY 347 -----AVNGNGLAPARGYADHTMG-----FPMSSVNG 375
Db 340 LGSRCGPRPGPRFEGNRPDPRPYEGHBAEGTKSKWGMIPRGPASQFVITSTLSLP 399
QY 376 MGRGO-----PEDAFGGRPHQCAWPQAPKQSPVILVQPGADPSGHQ-----LNID 421
Db 400 ROSGPQWKGKPAF--GQHQOQPKSQAEPLSGN-KEFLADTSSNQKNFKMQSAAFSIA 456
QY 422 RNIEGVK-----QPVVQDKPRVPMFGHCHPGELDW-TSMFOQAPQEGVMSQSMGP 472
Db 457 ADVKDVKAQAQENENLSDSQEPFKSEVSGPVPSNWDQNVQSMETQDKAQVTPVPL 516
QY 473 GOEPHAEVETKRYPTTTAGQESGMNGLVSLATMSGDGTVP 516
Db 517 ANKPVPA-----QSTFPBKTGMGGT--AVATSSLTADNDFKP 553

RESULT 9

ADCC31187
ID ADC31187 standard; protein; 2062 AA.
XX
AC ADC31187;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel polypeptide sequence, SEQ ID NO:1269.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 14q24.3.
XX
OS Homo sapiens.
XX
WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
FA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
XX WPI; 2003-371981/35.
DR N-PSDB; ADC30216.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 20; SEQ ID NO 1269; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human polypeptide sequence of the invention. Note: The sequence

Db 1093 HHGSAAP--PPTSAHQSPWFCSTSYICSSITVLLHLLHLLINHHGSAAP-----PPTSAHH 1146

Qy 393 G-APWQAPKQSP 404

Db 1147 GSAPPPTSAAHQSP 1159

RESULT 11

ID ADC07968

XX ADC07968 standard; protein; 1142 AA.

AC

XX

XX ADC07968;

DT 18-DEC-2003 (first entry)

XX

DE Rice protein sequence Seq ID234 related to grain filling.

XX

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;

KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;

KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;

KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;

KW gene; plant.

XX

OS Oryza sativa.

XX

PN WO2003000905-A2.

XX

PD 03-JAN-2003.

XX

XX 21-JUN-2002; 2002WO-IB002450.

PF

XX 22-JUN-2001; 2001US-0300112P.

PR

XX 26-SEP-2001; 2001US-0345227P.

PR

XX 20-DEC-2001; 2001US-0342327P.

XX

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PA

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;

PI

XX WPI; 2003-229341/22.

DR

DR N-PSDB; ABC07967.

XX

XX New plant genes encoding polypeptides having an activity involved in or

PT associated with the synthesis, metabolism or degradation of carbohydrates

PT in the plant grain useful in generating plants having improved

PT nutritional properties.

XX

XX Claim 15; SEQ ID NO 234; 130pp; English.

XX

CC This invention, in the area of plant biotechnology, relates to novel

CC polynucleotides comprising a nucleotide sequence encoding a protein which

CC is involved in or associated with the synthesis, metabolism or

CC degradation of carbohydrates in the plant grain and the expression of

CC which is up-regulated during grain filling. The plant is selected from

CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,

CC sugarbeet, wheat, and rice. The invention may be useful for the

CC improvement of protein, oil, starch, fibre and moisture content of the

CC cereal grains. In addition, carbohydrate levels may be modified to a more

CC desirable level using the present invention. The present sequence is the

CC amino acid sequence of a rice protein of the invention. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/publishedpct_sequences.

XX

SQ Sequence 1142 AA;

Query Match 8.3%; Score 238.5; DB 7; Length 1142;

Best Local Similarity 23.2%; Pred. No. 4.6e-08;

Matches 145; Conservative 56; Mismatches 243; Indels 181; Gaps 30;

Qy 3 PRNHP-----SRPPTSLPQGSAPLPSPAPISSMPMPQYTMQPYVSPHTLPPLQPHHS 57

Db 248 PQNQPHLTQLCAPAAQSQSQHPPLOSAQAQIQAQPSQSVSQPQPHAQPT-----T 302

Qy 58 QSPAPHSYMGQPPYRPDLNRYPASSHDVYASAPIMPHTTVGLPPTSFISHNPQQA 117

Db 303 HTPVPTQLGSGQFPAMPPTQATFSEHHQVMAQQQOPQLQH-----LPOC---QHPNAQQQS 354

Qy 118 QAO--QSPHYPPPHSVLPASSA-----QSPQPIAPAPPRDRRDNNGI-----P 162

Db 355 YPQQAIVHQPPPMHQAQFQNSVHAIVGHQFSQP-QPAHQMQQAPLQRLSHVSCQOMP 413

Qy 163 SGAFSYSDGKPGQWDP-VAANGAARPYFGKQSPRTQVVGSQ-RRGILSVFGRAPPTVNG 220

Db 414 SAQHHAHAHTPQGOQPTWMAQGI-----QQTPOHQHVGHHAALPEIYASIPQAAP----- 464

Qy 221 VNGTGKNTTIPAKDADGKFPNCPNCKTYLHA-----KHLKRLHLLRHTGRPYNCVLCK 273

Db 465 -QGFFLNAPAPSQTQSGSYQQGMPSSQLMHAPLQSQGQFMQOHPHTHTSAGRSNMYVAPQ 523

Qy 274 DTFSR-----

Db 524 EQFQNSGGPVKGLQAGVMNQPPMRMASDNVGTSELHGAGSFGQSGSSSLKXKPTSESE 583

Qy 290 -SIRRGNTGATHLSHPNAHVKRQQQAAANPV-----KPVQD-----EVSST 331

Db 584 KSENATNGTGNTEVSGKNG-----SAESALVNPISLDGSDGSKGKGVDFSAWESNSH 639

Qy 332 VPPNGIPGTTYG--EGAVNGNL--APARPGYAD-----HOTMGFPMSSVNGMGRGQ-- 380

Db 640 DPDARGGKTRSGISNDLVKGSLLQAQAFQHPYGPDSMLPQHMRQFGHMPYMQGLPNQMRP 699

Qy 381 PEDAFP--GGRPHQAGAPWQAPKQSPYLVOGADPSGHLNIDRNLIEQVKQPVQDPKRPV 439

Db 700 PKHSFPENSRRPMMQPPFENAPR-----VPG--PNQNMQISQSI-RPDGAIVRPFMGAP 750

Qy 440 MPG-HPGHGPGELDTSMFQPAPE-----GYMFSQ----- 468

Db 751 MPGLH-----DSTVPPFAPYVGTKKNSYGVNGPHGGSRALFEGGFNNSQKHKS 801

Qy 469 -SMFGGQEPYIAHVETERYKYYPTTT 492

Db 802 AANFGRRNVSHKDFEDNMKQFFVPT 826

RESULT 12

ABB65293

ID ABB65293 standard; protein; 1192 AA.

XX

AC ABB65293;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22671.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

XX WC200171042-A2.

PN

XX 27-SEP-2001.

PD

XX 23-MAR-2001; 2001WO-US009231.

PF

XX 23-MAR-2000; 2000US-0191637P.

PR

XX 11-JUL-2000; 2000US-00614150.

PR

XX (PEKE) PE CORP NY.

EA

XX Vester JC, Adams M, Li FWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR

DR N-PSDB; ABL09396.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 22671; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1192 AA;

Query Match 8.3%; Score 237; DB 4; Length 1192;
 Best Local Similarity 23.5%; Pred. No. 6.1e-08;
 Matches 137; Conservative 34; Mismatches 155; Indels 256; Gaps 32;
 QY 10 PPSTSL---PQGAFLPSA-----PISMPMPQYTMOPQY---PVSQPH 48
 Db 8 PPTQFQQQFQFGAPPNSGWPQQOQLPQQQFPQQQLPQQOQQQYQGAFFTSAAS 67
 QY 49 LPPLQPHHSQSPAPH-----SYMGOPYRDLNR-----YPASSHDVYASSAP- 92
 Db 68 QVYLGNYQQQLATSMGSLVGVGANLKPPLPQCAPAAAAAPPPTGNGQNSNAAPP 127
 QY 93 -----IMHETTVGS-----LPPTSLSHNPQAQA-QQS-----PHY 126
 Db 128 PTNNNAAGAPPTQAGSYNGALPPSS-----TPQSVASGNGSLNSATLAGLPHMP 182
 QY 127 PPHSVLP-----PASSAQSYQPIAPAPRRDRADFNGLPSGAFSYSDGK- --QG 175
 Db 183 PPKAATGAAPGQPPTPAAGTSQP-FLPQPP-----LP-----GQPPFSG 223
 QY 176 WDPVAANGAAPPYQKGSPTQVVGSGQRRGILPSVGRATPTVNGVNGTKNVTI 235
 Db 224 QIPTSQAPSPFY-----GVPSSRPQQLPFGATPPTVTPG-----LPQQQ 266
 QY 236 DGKFPCCNCKTYLHAKHLKHLRHTGDRYCMVLCCKDTFSRSDILKRFQKCSIRGN 295
 Db 267 QGIPP-----SGHQLNIDRNIQVKQPVQDPRVMP--GHPGHPCELDWTS 271
 QY 296 FTGATHLSHPNAHVKRSQQAAANPVKPVQDEVSTVPPNGIPGTTYGEGAVNGGLAP 355
 Db 272 -----LQOPGI---PQQQPGFPQQPGLPLLSQGLPQP--PGAPYG-----AP 310
 QY 356 ARGYADHQTMGFPMSVNGMRGQPEDAFG-----GRPHQGAFWQAPKQSP 404
 Db 311 QQQGYSG-----GFP-----GQAPGFPFAPPLPQQAAPQFGAPQPGYQGPQ 357
 QY 405 -YLVPQGADP-----SGHQLNIDRNIQVKQPVQDPRVMP--GHPGHPCELDWTS 454
 Db 358 GYPPQPCQQMPGYPPQFGQQLG-----GPGYPPQFGAGFPQGPGRPGFN- 402
 QY 455 MFQQAQ-EGYMFQS-----MPGQEPHIAHVETER 485
 Db 403 -QPPMFGAGMYQQAQARRLDPDQMP---NPIQVMIENQR 439

RESULT 13
 ABB70436
 ID ABB70436 standard; protein; 644 AA.
 XX
 AC ABB70436;
 XX
 DT 26-MAR-2002 (first entry)

XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 38100.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL14539.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 38100; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 644 AA;

Query Match 8.3%; Score 236.5; DB 4; Length 644;
 Best Local Similarity 26.1%; Pred. No. 3.3e-08;
 Matches 111; Conservative 43; Mismatches 162; Indels 109; Gaps 20;
 QY 8 SRPSTSLPQGSAPLPSAPISMPMPQYTMOPQYVVSQPHLPPLOPHHSQSPAPHSYMG 67
 Db 100 SAPPSLSGSSGSSGSSPLYGKP-----PMKLELPYQASSTGTASPNSSIQAPSASV 155
 QY 68 QPYRDLNRYPASSHDVYAS--SAAPIMPTHTVGSLLPPTSFL-----SHPNPQ-- 114
 Db 156 SPISFPS-----FAQS-----FASISASPTTTTUA--PPTTAAAGALAGSFTSSSPSSAA 206
 QY 115 -----AQQAQQSQPHYPHPSVLPASSAQSYQPIAPAPRRRADF 157
 Db 207 SAAAAAATAAADAAGAAVASAAVGNWTAISGLGPARSQPFYAQ-----YASDY 257
 QY 158 NN---GLPSCA--FYSYDGKQFQWDVPAANGAAPYQKDSPTQVVGSGQRRGILSVPG 212
 Db 258 YGNAVGMSSSAAWFQHERLYQPW-----SSQSYPGFNDDIAFOTQLQRRSVRCPCN 311
 QY 213 RATPTVINGTGKNTTIIPAKDAGK----FPCPNCKTYLHAKHLKHLRHTGDRPYM 268
 Db 312 ----CTNEMSGL---PPIVGPDERGRKQHIHIPGCELYGKASHLTKHLRHTGDRPFL 364
 QY 269 CVLCKDTFSRSDILKRFQ-----KCSIRGNPTGATHLS-HPNAHVK----- 310
 Db 365 CLTCGKFRSRDELQHRGHTHNYRYPACFICKKFSRSDHLSKHKKTHFKOKSKKVL 424
 QY 311 -RSQQQAAA-----NPVKP---VQDEVSTVPPNGIP--GTYGEGAVNGNGL 353

Db 425 AEAKEQAAAAIKLEKKEKSGKPLTPPVVEFKQEQDPTTFLVNYAPYANLYQHSSTAGSSV 484
QY 354 APARP 358
Db 485 NPPPP 489

RESULT 14
AA333497
ID AA333497 standard; protein; 1185 AA.
AC AA333497;
DT 19-JAN-2000 (first entry)
DE Human atrophin I protein.
XX
KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCAl; SCAl; SCAl;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidolysian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic.
XX
OS Homo sapiens.
XX
XX WO9945944-A1.
XX
PD 16-SEP-1999.
XX
PF 11-MAR-1999; 99WO-US005250.
XX
PR 12-MAR-1998; 98US-00041886.
XX
FA (BURN-) BURNHAM INST.
XX
PI Bredesen DE, Rabizadeh S;
XX
XX WPI; 1999-561617/47.
DR N-PSDB; AA223430.
XX
XX
PT New proapoptotic dependence peptides, used to develop products for
PT treating, e.g. Alzheimer's disease.
XX
PS Disclosure; Page 154-157; 199pp; English.
XX

This invention describes novel pure proapoptotic dependence peptides
which comprise a sequence of an active dependence domain selected from
dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
huntingtin polypeptide, Machado-Joseph disease gene product, SCAl, SCAl,
SCAl and atrophin-1 polypeptide. The proapoptotic peptides are capable of
inducing cell death and can be used to develop products to mediate or
inhibit apoptosis. The methods can be used for reducing the severity of a
proapoptotic dependence domain mediated pathological conditions e.g.
Huntington's disease, Alzheimer's disease, Kennedy's disease,
Spinocerebellar ataxia, dentatorubropallidolysian atrophy, Machado-
Joseph disease, stroke or head trauma. They can also be used for reducing
the severity of a pathological condition mediated by upregulated cell
proliferation or cell survival e.g. neoplastic, malignant, autoimmune or
fibrotic conditions. This sequence represents the human atrophin I
polypeptide described in the method of the invention

Query Match
Best Local Similarity 8.1%; Score 233.5; DB 2; Length 1185;
Matches 123; Conservative 46; Mismatches 193; Indels 197; Gaps 23;
QY 7 PSRPPSTLPGSAPLPSAPITSPMPQYTWQPYVPSQPTLPLPQPHHSQSPAPHSYM 66
Db 272 PTKPPTT--PVGGGNLPSAP-----PPANFPHVTNLP----- 303
QY 67 QPPYRPLNRYPASSHDVYASSAAPIMPHVTGSLPSTSLSHNQAQA----- 119

Db 304 --PPALRPLNNASPPGL--GAQLFGLH-----PSPYAMQMGGLPG 345
QY 120 -QQSPHY-PPSHSVLPASSAQSYQPIAPAPPRDRRADFNGLPSGAFSYSDGKPGQMD 177
Db 346 PEKGTLPSPHS-LPPASS-----APAPPMFPYSSSSSSSSSSSSSSSSS 397
QY 178 PVAANGAAP-YPGKDSPTQV-VGSQRRGILPSVGRATPVTVNGVGTGKNTTIPAKDA 235
Db 398 PFPASQALPSYHSPPTSLSVSNQPKYQPSLPSQAV-----WS 439
QY 236 DGKPEPCNCKNTYLHAKHLKRLHHTGDRPYMVCVLCXDTFSRSDILKHFQKCSIRGN 295
Db 440 QGPPPPPPYGRLLANS-----AHPGFP-----P 464
QY 296 PTGATHLSHP---NAHVKRQQQAAANFVKPVQDEVSVTPPPNGIPGTTYGEGAVNGN- 351
Db 465 STGAQSTAHPPVSTHHHHHHQQQQQQQQQHHGNSGPPPPGAPPHLEGSSHHAP 524
QY 352 -----GLAPARPG-----YADHTMGFFMSSVNGMGRQPEDAFPGGRPH 391
Db 525 YAMSPSLGSLRPYPFGPAHLPPHPSQVSYSQAGNPPVSSSSSSSSSSSSSTQSGYPCSHPS 584
QY 392 -----QGAPWQAP-----KQSPYLVOGADPSGHOLNIDRNIQVQPVVQDPCR 437
Db 585 PSQGPQAGPYFPFVPTVTTSSATLSTVIATVASPAGY-----KT 625
QY 438 PVMPGHPGHGELDTSMFQPAPEGYMFQSGMFGQPIHARVETERKYPTTTAGORS 497
Db 626 ASPGPPPYGKRAPSPGAYKTATPGY-----KFGSP-----PSFRTGTPP 666
QY 498 GMNGLYLASTMSGDGTQVP 516
Db 667 GYRG---TSPPAGEPTFKP 682

RESULT 15
ABB65135
ID ABB65135 standard; protein; 926 AA.
XX
AC ABB65135;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 22197.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL09238.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 22197; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (AB857737-
CC AB872072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 926 AA;

```
Query Match      8.0%; Score 229.5; DB 4; Length 926;
Best Local Similarity 22.9%; Pred. No. 1.5e-07;
Matches 142; Conservative 44; Mismatches 229; Indels 205; Gaps 30;

Qy 4 RNHP-----SRPPTSLLP-QGSAPLPS-----AFISSMMPQYTMQP 39
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
351 QQHPVYRNAQGGQGGAGQVFGQGGPVQSVINENAAFNQORPNNGLSLGPNFQOQQQQ 410
Qy 40 QYVVSQHTLPPQLPHHSQSAPHSYMGQPYRDLNRYPASSHDVYASSAAPIMPHITV 99
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
411 PQGGQQ---PPNQOQQOQQQTGP-----GGPGQP-----GAGGCGVP--- 445
Qy 100 GSLPPTGFLSHPNPQAQAQAQOQSHYPP-----PHSVLPASSAQSYPQPIAPAPPRD 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
446 ---PPQSPYRVSYQQOQQOQSHYCYPPQPTQYQPOQAYPGPTQGYGPP-PPGPNNA 501
Qy 153 RRADFNGLPSGAFSYSGKPGQMDPVAANGAAPYPKDSPTQVVGSGRRGILPSVPG 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
502 AQGGYHHG-PAGAAATGASG--HGYPNAGAGCGPPPGAYPPPP---GSQQ-----VPEVPG 551
Qy 213 RATEVTN----GVNGTGKNTIPAKDADGKE-PCPNCNKTYLHAKHLKRHLRHHTGDRPY 267
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
552 QQPPPPPPPPGPPPTGGQQOQQPPGPPQSQSYGPPPPQNS-----AGGPPP 596
Qy 268 MCVLCKDTFSRSDILKRHFQKCSIRRGNTG-----ATHLSHPNAHVKRSQQQAAA 318
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 M-----GYAGYPNPQYGGAGAGGPPPSGYWFPPTSSAQSYPYQYQQOQQOQAAA 650
Qy 319 N-----GVKPYDEVST-----VPPNGIPGITYGEGAVN----- 349
Db 651 GGGAGAPGSSYPGGRPTSGAAPPPPPGAYSTTAPSTPPQGGGAGGNNPNGPNA 710
Qy 350 -----GNGLAPARPGYADHQTWGFPMSSVNGMRGQPEDAPPGGRPHQG-APWP 397
Db 711 QOSTPPQGGAGGAGSGPGGAGQQYAGPPQQ-----QPQQQPPGVVYSGVAPLP 762
Qy 398 ---QAPKQSP--YLVQPGADPSGHQLNIDRNIHQVKQPVV----- 432
Db 763 TVQQTYSTGNYNQFGAPPNPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 822
Qy 433 -QDP-----KRPVMPGCHPGCELDWTSNFOQPAPEG-----YMFQSMPGQCEPIHA 479
Db 823 NQPPPNNGATPMPMPNQYQPA-----FGAPQGYGPPPPPPQAYGPPPPGSAYPEGHA 873
Qy 480 HVETER---KYYPYTTIAGQ 495
Db 874 YHQPQAGGYAQYPPYQGYQ 893
```

Search completed: April 1, 2004, 17:46:52
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:39:33 ; Search time 17 Seconds
(without alignments)
1589.671 Million cell updates/sec

Title: US-10-029-180-8

Perfect score: 2866

Sequence: 1 MDPNRHSPRSTSLPQSA.....NGLYLASTMSDGTVPQPARQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	255.5	8.9	722	1	Z219 HUMAN
2	251.5	8.8	2716	1	Q9p2y4 homo sapien
3	236.5	8.3	644	1	O8in94 drosophila
4	234.5	8.2	1185	1	P54266 drosophila
5	233	8.1	856	1	P54259 homo sapien
6	225	7.9	692	1	Q60636 mus musculus
7	223	7.8	470	1	O74252 emericella
8	222	7.7	963	1	P08152 mus musculus
9	221.5	7.7	431	1	Q8tdd2 caenorhabdi
10	221.5	7.7	582	1	Q10076 schizosach
11	221	7.7	470	1	P51774 rattus norv
12	218.5	7.6	5262	1	O14686 homo sapien
13	212.5	7.4	1183	1	P54258 rattus norv
14	211.5	7.4	354	1	Q60843 mus musculus
15	211	7.4	789	1	O75526 homo sapien
16	210.5	7.3	543	1	P18146 homo sapien
17	209.5	7.3	533	1	P08046 mus musculus
18	209	7.3	398	1	Q91hxx2 mus musculus
19	208.5	7.3	323	1	Q8td94 homo sapien
20	207.5	7.2	1386	1	Q9r0i7 mus musculus
21	207	7.2	476	1	P11161 homo sapien
22	206.5	7.2	508	1	P08154 rattus norv
23	206.5	7.2	5120	1	Q9pu36 gallus gall
24	206	7.2	428	1	Q8vi67 mus musculus
25	206	7.2	1157	1	O95104 homo sapien
26	205.5	7.2	345	1	P57682 homo sapien
27	205.5	7.2	5147	1	Q9y6v0 homo sapien
28	202	7.0	1822	1	P49750 homo sapien
29	201.5	7.0	1300	1	Q9bxa9 homo sapien
30	201.5	7.0	5085	1	Q9jks6 rattus norv
31	200.5	7.0	446	1	Q9z0z7 mus musculus
32	200	7.0	362	1	Q13351 homo sapien
33	199.5	7.0	481	1	Q9nw07 homo sapien

34	199.5	7.0	511	1	EGR1_BRARE
35	198.5	6.9	449	1	WT1_FIG
36	198.5	6.9	4903	1	MLL3_MOUSE
37	198	6.9	344	1	KLF3_MOUSE
38	198	6.9	351	1	KLF2_RAT
39	197.5	6.9	421	1	EGR2_XENLA
40	197.5	6.9	780	1	YNL5_CAEEL
41	197	6.9	432	1	BRLA_EMENT
42	196.5	6.9	477	1	MAZ_HUMAN
43	195.5	6.8	449	1	WT1_HUMAN
44	194.5	6.8	467	1	CBP4_DICDI
45	194.5	6.8	773	1	Z341_HUMAN

P26632 brachydanio
O62651 sus scrofa
Q8brh4 mus musculus
Q60980 mus musculus
Q9et58 rattus norv
Q08427 xenopus lae
Q21955 caenorhabdi
P10069 emericella
P56270 homo sapien
P19544 homo sapien
P35085 dictyosteli
Q9byn7 homo sapien

ALIGNMENTS

RESULT 1
Z219_HUMAN
ID Z219_HUMAN STANDARD; PRT; 722 AA.
AC Q9P2Y4; Q9BW28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 219.
GN ZNF219.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=20277481; PubMed=10819330;
RA Sakai T., Toyoda A., Hashimoto K., Maeda H.;
RT "Isolation and characterization of a novel zinc finger gene, ZNF219,
RT and mapping to the human chromosome 14q11 region.";
RL DNA Res. 7:137-141 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan F.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR ENBL; AB015427; BAA90526.1; -;
 DR ENBL; BC000694; AAH00694.1; -;
 DR Genew; HGNC:113011; ZNF219.
 DR MIM; 605036; -;
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0006351; P:transcription, DNA-dependent; TAS.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 9.
 DR SMART; SM00355; Znf_C2H2; 9.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
 DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT ZN_FING 57 79 C2H2-TYPE.
 FT ZN_FING 85 107 C2H2-TYPE.
 FT ZN_FING 163 186 C2H2-TYPE.
 FT ZN_FING 274 296 C2H2-TYPE.
 FT ZN_FING 302 324 C2H2-TYPE.
 FT ZN_FING 498 520 C2H2-TYPE.
 FT ZN_FING 520 520 C2H2-TYPE.
 FT CONFLICT 232 233 MISSING (IN REF. 2).
 FT CONFLICT 436 436 Q -> E (IN REF. 2).
 SQ SEQUENCE 722 AA; 76876 MW; B19DA77B14B8C45B CRC64;

Query Match 8.9%; Score 255.5; DB 1; Length 722;
 Best Local Similarity 24.38; Pred. No. 3.7e-06;
 Matches 145; Conservative 46; Mismatches 156; Indels 249; Gaps 35;

QY 7 PSRPTSLPQGSAPLPSAPISMPQYTMQPPVVSQHTLPPQLPHHSQSPAP-HSY 65
 Db 215 PERPLAAT--SAAPP-----QPQPPQPE-PRSVQPEPEPEPEEPTTPAPAA 264
 QY 66 MGQPPYPD-----LNYPAS-----SH-DVY 86
 Db 265 PEEPPAPPERPCVCGQSQTSQWFLKGHRKHKASFDHACVCGRCFKEPWFLLKMKVH 324
 QY 87 ASSAAMPHTTSGLSFTSLFSLHNP-QAQAQSQSPH-----YPPHVSPLPASSAQ 140
 Db 325 ASKLG-----LRAPGASGAPAPQPDGLIAYE-----LGPALL--- 362
 QY 141 YPOPIAPAP-PRDRADFNGLPSGAFSY-----SDGKPOQWDVAANGAAPPYPKD--- 191
 Db 363 ----LAPAPTAERRE-----PSSLGLYLSRAGEGRNG-----EGAEPPGGRSGFG 406
 QY 192 -SPRTQVVGSGRR-----GILPSV-----PGR----- 213
 Db 407 FRPLSSALPARARHRAEPEEVEVVEAQPETWARGSLGSLASLHPRFPGPGHGSASA 466
 QY 214 -----ATPVNG--VNGT---GKNTTIPAKDAGKPCPCNKNTYLHAKLKHLL 259
 Db 467 AGAARATATQENGLLVGTRPREGGAT--GKD-----CPGKSPFSAHLKRVHLR 518
 QY 260 RHTGDREYMCVLKXDTSRDLKRRPKQSGIRRGNTGTATHLSHPNAHKRS---QQQA 316
 Db 519 VHTGERYPKPCDYAGTQSGSLKYHLQRHREQRSGAGPGPPEPPPPSQSGAPQSGA 578
 QY 317 AANPKVQDVSVSTVPPNGIPGTTTGEQAVNGNGLAPRGVADHQTGFMFMSVNGM 376
 Db 579 KPSQPATWVEGASSPRPS-----SGAGPGSRKPKASPG-----RTLR-----NGR 620
 QY 377 GRGQPE-----DAFFGGRPHQ-----GAPW-----P 397
 Db 621 G-GEAEFLDLSLRAGPGEAGPGGALHRCILFCPPATGAPELMALHLQVHHSRRARRPP 679
 QY 398 QAPKQSYLVQPGADPPSGHQLNIDRIEIQVKQPVVQDPKPKPVMEHPFG-----HPGE 449
 Db 680 QADASPFY---ARVPSG-----ETPPSPSQGBEGSGLSRPE 714

RESULT 2

OSA_DROME
 ID OSA_DROME STANDARD; PRT; 2716 AA.
 AC Q81N94; O61603; Q9VEG7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Trithorax group protein OSA (Eyelid protein).
 GN OSA OR ELD OR CG7467.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
 RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=97415319; PubMed=9271118;
 RA Treisman J.B., Luk A., Rubin G.M., Heberlein U.;
 RT "eyelid antagonizes wingless signaling during Drosophila development
 RT and has homology to the Bright family of DNA-binding proteins.";
 RL Genes Dev. 11:1949-1962(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cwley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira J.A., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP FUNCTION AND DEVELOPMENTAL STAGE.
 RX MEDLINE=99112962; PubMed=9895321;
 RA Vazquez M., Moore L., Kennison J.A.;
 RT "The trithorax group gene osa encodes an ARID-domain protein that
 RT genetically interacts with the brahma chromatin-remodeling factor to
 RT regulate transcription.";
 RL Development 126:733-742(1999).
 RN [4]
 RP DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.

EX MEDLINE=20069333; PubMed=10601025;
RA Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RT "Osa associates with the Brahma chromatin remodeling complex and
RL promotes the activation of some target genes.";
RN EMBO J. 18:7029-7040(1999).
[5]
RP FUNCTION.
RX MEDLINE=99403006; PubMed=10471712;
RA Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
RT "A genetic screen for modifiers of E2F in Drosophila melanogaster.";
RL Genetics 153:275-287(1999).
[6]
RP IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RX BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RA MEDLINE=20270023; PubMed=10809665;
RX Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
RT "The Drosophila brahma complex is an essential coactivator for the
RL trithorax group protein zeste.";
RN Genes Dev. 14:1058-1071(2000).
[7]
RP FUNCTION AS A COREPRESSOR.
RX MEDLINE=20573925; PubMed=11124806;
RA Collins R.T., Treisman J.E.;
RT "Osa-containing Brahma chromatin remodeling complexes are required
RL for the repression of wingless target genes.";
RN Genes Dev. 14:3140-3152(2000).
[8]
RP FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
RX MEDLINE=22515897; PubMed=12629041;
RA Heitzler P., Vanclost L., Biryukova I., Raman P.;
RT "Enhancer-promoter communication mediated by Chip during
RL panier-driven proneural patterning is regulated by Osa.";
RN Genes Dev. 17:591-596(2003).
CC -!- FUNCTION: Trithorax group (trxG) protein required for embryonic
CC segmentation, development of the notum and wing margin, and
CC photoreceptor differentiation. Required for the activation of
CC genes such as Antp, Ubx and Eve. Binds to DNA without specific
CC affinity, suggesting that it is recruited to promoters by
CC promoter-specific proteins. Essential component of the Brahma
CC complex, a multiprotein complex which is the equivalent of the
CC yeast SWI/SNF complex and acts by remodelling the chromatin by
CC catalyzing an ATP-dependent alteration in the structure of
CC nucleosomal DNA. This complex can both serve as a transcriptional
CC coactivator or corepressor, depending on the context. Acts as an
CC essential coactivator for Zeste, which recruits the whole complex
CC to specific genes. In contrast, it acts as a corepressor for Wg
CC target genes, possibly via an interaction with Pan and Gro. It
CC also acts as a negative regulator for proneural achaete-scute,
CC when it is directly recruited by Pan and Chi. Also represses E2f
CC activation.
CC -!- SUBUNIT: Component of the Brahma complex, which is composed of
CC Brm, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
CC Interacts with Pnr and Chi via its EHD domain.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
CC third instar larvae, it is ubiquitously expressed in wing and eye-
CC antenna imaginal disks, with a stronger expression in a band just
CC anterior to the morphogenetic furrow.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC -!- DOMAIN: The ARID domains mediates the binding to DNA.
CC -!- SIMILARITY: Contains 1 ARID domain.
CC -!- SIMILARITY: Contains 1 EHD (Yelid homology) domain.
CC -!- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF053031; AAC06254.1.1; -;
DR EMBL; AE003718; AAP55457.1; ALT_SEQ.
DR EMBL; AE003718; AAN13750.1; -;
DR PIR; T13049; T13049.
DR FLYBase; FBgn0003013; osa.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
DR GO; GO:0045449; P:regulation of transcription; IDA.
DR GO; GO:0007379; P:segment specification; IMP.
DR GO; GO:0008587; P:wing margin morphogenesis; IMP.
DR GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR006031; XYPPX.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02162; XYPPX; 8.
DR SMART; SM00501; BRIGHT; 1.
KW Transcription regulation; DNA-binding; Activator; Repressor;
KW Chromatin regulator; Nuclear protein; Developmental protein.
FT DOMAIN 997 1111
FT DOMAIN 1769 2517 EHD.
FT DOMAIN 19 1763 PRO-RICH.
FT DOMAIN 174 380 GLN-RICH.
FT DOMAIN 619 873 GLY-RICH.
FT DOMAIN 1222 1453 GLY-RICH.
FT DOMAIN 1271 1751 GLN-RICH.
FT DOMAIN 1730 1745 HIS-RICH.
FT DOMAIN 2589 2624 SER-RICH.
FT DOMAIN 2625 2716 ALA-RICH.
FT CONFLICT 61 61 MISSING (IN REF. 1).
FT CONFLICT 1169 1169 V -> G (IN REF. 1).
FT CONFLICT 1795 1795 M -> T (IN REF. 1).
FT CONFLICT 2637 2637 G -> E (IN REF. 1).
SQ SEQUENCE 2716 AA; 284063 MW; EFAB76CB51C7C675 CRC64;
Query Match 8.8%; Score 251.5; DB 1; Length 2716;
Best Local Similarity 23.9%; Pred. No. 2.1e-05;
Matches 152; Conservative 49; Mismatches 208; Indels 227; Gaps 36;
QY 3 PRNPSR----PPSTSLPQGSAPLESA-----PISSMPMPQVTMPQVPSQ----- 45
DB 292 POLGFSQQYRTPTNTSRGQSPVPPAHGQNSGVPSPPSQQQQQQQQQQQAGQPGP 351
QY 46 -PHILPP---LOPHSQSPASHVMGPPYPDLNRYPA--SSH-DVTASSAAPIMPTT 98
DB 352 VPGGPPPTGTGQQPQQNTPTTSQSPYQRYTPTGPGAGSNHRTAYSTHQYPPNRPW 411
QY 99 VGSLLPPTSLSHPNPQAQAQSQSPHY-----PPHSVL--PPASSAQSY-----P 142
DB 412 PGGSPSPGSGHPLPPA-----SPHHVPPLOQQPPPPPHYSAGGPPSSSGHAPSPSP 465
QY 143 QP--IAPAPRRDRADFNGLPFGAFYSYDQKPGQDWDVAANGAAPYPGKDSPTQVY-- 198
DB 466 QPSQASPSFHQLICQNSDSSSG-----GAHSGM-----GSGP-PGTENPQ-QVMRP 511
QY 199 -----GSGRGRLPSYPGR---ATPYTNGVNGTKNTTIKADADGKFCPCNCKTYLH 250
DB 512 TFSPTGSGSGSRSMGFAVNAQNHPISRPNASN-QSSSGGPMQPPVAGAGGPPMP----- 562
QY 251 AKHLKRLRLHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSRRRGNTGATLHSHPNHVK 310
DB 563 -----PH-----PGMPGP-----PQQQ 575
QY 311 RSQQQAAN-----PVKPVQDEVSTVPPPPNGIPGTTY-----GEG 346
DB 576 QSQQQASNSASSASNSPQTPPPAPPNNQGMNNMATPPPPQGAAGGYPMPHMHGY 635
QY 347 AVNGGLAPAPPGYADHTMGF-----PMSSVNGMG--RGOPE 382
DB 636 KMGFGSPGAGQYPPQPPQYPPGNTVPPRPYPPGAYATGPPPTTQAGAGGANSNPS 695
QY 383 DAFFGGRPHQGP-----W-POAPKQSPYLQVFGADFSGHQLNIDRNIHQV--- 428

```

Db 696 GAQAGGPGGMPNHTGQYPPYQWPPSPQOT-----VPGAPGAMVG-----NHVQGGTTP 748
Qy 429 -QPVVQPKRPVMPGHP-----GHPGLDWTSMFQPAPEGY-----MFSQ 468
Db 749 PPPVGGPPPPQSGSRPLNLYKHQHLQKGGYGG-----SPTPPQPGQYGGNGTGMHP 803
Qy 469 SMPGQGPPIH---AHVTERKYYPTTTAGBSGNMG 501
Db 804 GMPMG-PPHMGPPHGGTNGM-PTSTPPQSQMLQG 837

RESULT 3
BTID_DROME
AC Q24266; Q3W319; STANDARD; PRT; 644 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor BTID (Buttonhead protein).
GN BTID OR CGI2653.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=94081952; PubMed=8259212;
RA Wimmer E.A., Jacek H., Pfeifle C., Cohen S.M.;
RT "A Drosophila homologue of human Sp1 is a head-specific segmentation
RL gene";
RN Nature 366:690-694 (1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RP
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen C.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
CC -!- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF THE ANTENNA,
CC INTERCALARY AND MANDIBULAR SEGMENTS OF THE HEAD.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN A STRIPE COVERING THE HEAD
CC ANLAGEN OF THE SYNCHIAL BLASTODERM EMERVO, PERSISTS THROUGH
CC GASTRULATION AND DECAYS DURING GERM BAND EXTENSION. EXPRESSED
CC LATER IN DEVELOPMENT IN A COMPLEX SPATIALLY RESTRICTED PATTERN.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z29361; CAA82545.1; --
CC EMBL; AE003448; AAF46518.1; --
CC HSP; P08047; ISP2.
CC FlyBase; FBgn0000233; btd.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.
CC ProDom; PD000003; Znf_C2H2; 1.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 333 357 C2H2-TYPE 1.
FT ZN_FING 363 385 C2H2-TYPE 2.
FT ZN_FING 391 413 C2H2-TYPE 3.
FT DOMAIN 14 89 GLN-RICH.
FT DOMAIN 208 220 POLY-ALA.
FT DOMAIN 431 434 POLY-ALA.
FT DOMAIN 486 492 POLY-PRO.
FT DOMAIN 499 502 POLY-THR.
FT DOMAIN 515 519 POLY-SER.
FT DOMAIN 530 536 POLY-SER.
FT DOMAIN 596 599 POLY-SER.
SQ SEQUENCE 644 AA; 68581 MW; A0DB98C2AF938452 CRC64;
Query Match 8.3%; Score 236.5; DB 1; Length 644;
Best Local Similarity 26.1%; Pred. No. 2.9e-05;
Matches 111; Conservative 43; Mismatches 162; Indels 109; Gaps 20;
Qy 8 SRPSTSLPGSAPLPSAPISMMPMQYTMQPVYVQSPHTLPPLQPHSQSPAPHSYMG 67
Db 100 SAPPSLSSGSSGSSSPLYGKP-----PMKLELPYQASSTGTASPNSIQAPSSASV 155
Qy 68 QPYRDLNRYPASHSDVYAS--SAAPIMPHITVGSPLPSTFL-----SHPNPQ-- 114
Db 156 SPISFIS---PAQS---FASISASPSPTTTTLA--PPTTAAAGALAGSPTSSSPSSAA 206
Qy 115 -----AQQAQSQSHYPHPHPSVLPASSAQSPQPIAPAPDRREADF 157
Db 207 SAAAAAAGAAAAADGAAVAAGYNTAYSGLGPARSQFFYQAQ-----YASDY 257
Qy 158 NN---GLPSCA---FSYSDGKPGQWDVPAANGAAYPGKQSPRTQVVGSGQGRGILPSVPG 212
Db 258 YGNAVGMSSSAAAFSHOERLYQW-----SSSYGFGNFDDIAFQTLQRRSVRCTPN 311
Qy 213 RATEVTNGVNGTKNTTIPAKDAGK-----FPCPNCKNTLYLHAKHLKRHLRHTGDRPYM 268
Db 312 ----CTNMSGSL---PPIVGPDERGRKQICHIPGGERLYGKASHLTKHLRHWITGERPPL 364
Qy 269 CVICKDTFSSDILKHFQ-----KCSIRGNPTGATHLS-HPNAHVK----- 310
Db 365 CLTCGKFRSRSDELQRHGRHTHTNYPYACPICSKFKFSRDLHSLKHKTKHFKDKSKKVLVA 424

```

QY 311 -RSQQQAAA-----NPVKP---VQDEVSTVPPNGIP-GTYYGEGAYNGEL 353
 Db 425 ABAKEQAAAIAKLEKKEKSGPLPPVFEKQBPQPTTJLVNVPYANLYQHSSTAGSSV 484
 QY 354 APARP 358
 Db 485 NPPPP 489

RESULT 4
 DRPL HUMAN
 ID DRPL HUMAN STANDARD; PRT; 1185 AA.
 AC P54259; Q99495; Q99621; Q9UBK7;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
 GN DRPLA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Cerebellum;
 RX MEDLINE=95144175; PubMed=7842016;
 RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
 RA Inoue T., Yamada M.;
 RT "Structure and expression of the gene responsible for the triplet
 RT repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";
 RL Nat. Genet. 8:177-182(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96026098; PubMed=7485154;
 RA Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Tsuji S.;
 RT "Molecular cloning of a full-length cDNA for dentatorubral-
 RT pallidolysian atrophy and regional expressions of the expand alleles
 RT in the CNS.";
 RL Am. J. Hum. Genet. 57:1050-1060(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=9622314; PubMed=8965642;
 RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
 RA Kidwai A.S., Ashworth R.G., Ross C.A.;
 RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
 RT brain.";
 RL Brain Res. Mol. Brain Res. 36:219-226(1996).
 RN [4]
 RP SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.
 RC TISSUE=Brain;
 RX MEDLINE=97228904; PubMed=9074930;
 RA Anzari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;
 RT "Large-scale sequencing in human chromosome 12p13: experimental and
 RT computational gene structure determination.";
 RL Genome Res. 7:268-280(1997).
 RN [5]
 RP SEQUENCE OF 1-76 FROM N.A.
 RX MEDLINE=97005364; PubMed=8852663;
 RA Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y.,
 RA Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K.,
 RA Soerensen S.A., Potter N.T., Young S.R., Nakamura K., Nukina N.,
 RA Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T.,
 RA Kanazawa I., Yamada M.;
 RT "A unique origin and multistep process for the generation of expanded
 RT DRPLA triplet repeats.";
 RL Hum. Mol. Genet. 5:373-379(1996).
 RN [6]
 RP SEQUENCE OF 470-725 FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=93315145; PubMed=8325628;
 RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
 RT "Novel triplet repeat containing genes in human brain: cloning,

RT expression, and length polymorphisms.";

RL Genomics 16:572-579(1993).
 RN [7]
 RP INTERACTION WITH WWP1 AND WWP2.
 RX MEDLINE=98313405; PubMed=9647693;
 RA Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan K., Kushi J.,
 RA Kaminsky Z., Kleiderlein J.J. Jr., Sharp A.H., Ross C.A.;
 RT "Atrophin-1, the DRPLA gene product, interacts with two families of
 RT WW domain-containing proteins.";
 RL Mol. Cell. Neurosci. 11:149-160(1998).
 CC -!- SUBUNIT: Interacts with WWP1 and WWP2.
 CC -!- TISSUE SPECIFICITY: Relatively high levels in the brain, ovary,
 CC testis and prostate. Lower levels in the liver, thymus and
 CC leukocytes.
 CC -!- POLYMORPHISM: The poly-Gln region of DRPLA is highly polymorphic
 CC (7 to 23 repeats) in the normal population and is expanded to
 CC about 49-75 repeats in DRPLA patients. Longer expansions result in
 CC earlier onset and more severe clinical manifestations of the
 CC disease.
 CC -!- DISEASE: Defects in DRPLA are the cause of dentatorubral-
 CC pallidolysian atrophy (DRPLA) [MIM:125370], an autosomal dominant
 CC neurodegenerative disorder characterized by a loss of neurons in
 CC the dentate nucleus, rubrum, globus pallidus and Luys' body.
 CC Clinical features are myoclonus, epilepsy, dementia, and cerebellar
 CC ataxia. Onset of the disease occurs usually in the second decade
 CC of life and death in the fourth.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to several
 CC frameshifts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31840; BAA06626.1; -;
 DR EMBL; D38529; BAA07534.1; ALT_FRAME.
 DR EMBL; U23851; AAB50276.1; -;
 DR EMBL; U47924; AAB51321.1; -;
 DR EMBL; D63808; BAA23631.1; -;
 DR EMBL; L10377; -; NOT_ANNOTATED_CDS.
 DR PIR; G01783; G01763.
 DR Genew; HGNC:3033; DRPLA.
 DR MIM; 607462; -;
 DR MIM; 125370; -;
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007417; P:central nervous system development; TAS.
 DR InterPro; IPR002951; Atrophin.
 DR Pfam; PF03154; Atrophin-1; 2.
 DR PRINTS; PR01222; ATROPHIN.
 KW Triplet repeat expansion; Polymorphism; Epilepsy.
 FT DOMAIN 73 82
 FT SER/GLU-RICH (MIXED CHARGE).
 FT POLY-PRO.
 FT DOMAIN 302 305
 FT POLY-SER.
 FT DOMAIN 376 382
 FT POLY-SER.
 FT DOMAIN 386 397
 FT POLY-SER.
 FT DOMAIN 442 447
 FT POLY-PRO.
 FT DOMAIN 479 483
 FT POLY-HIS.
 FT DOMAIN 484 497
 FT POLY-GLN.
 FT DOMAIN 504 507
 FT POLY-PRO.
 FT DOMAIN 564 574
 FT POLY-SER.
 FT DOMAIN 704 707
 FT POLY-PRO.
 FT DOMAIN 802 815
 FT ARG/ALA-RICH (MIXED CHARGE).
 FT DOMAIN 816 827
 FT ARG/GLU-RICH (MIXED CHARGE).
 FT DOMAIN 925 934
 FT MISSING (IN REF. 3).
 FT DOMAIN 94 94
 FT CONFLICT
 FT H -> Y (IN REF. 1).
 FT CONFLICT 333 333
 FT M -> I (IN REF. 3).
 FT CONFLICT 339 339
 FT P -> T (IN REF. 6).
 FT CONFLICT 541 541
 FT CONFLICT 1028 1028
 FT A -> G (IN REF. 1).

```
SQ SEQUENCE 1185 AA; 124773 MW; 5135744CEE491C18 CRC64;
Query Match 8.2%; Score 234.5; DB 1; Length 1185;
Best Local Similarity 22.0%; Pred. No. 6.5e-05;
Matches 123; Conservative 46; Mismatches 193; Indels 197; Gaps 23;

QY 7 PSRPPSTLPGSAPLPSAIPSSMPQYTMQPVQVQPHTLPLPQHHSQSPAPHSYM 66
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 67 GQPPYRDINRYPASSHDVYASSAAPIMPHITVGLSPPTSLFSDNPAQAQA-----119
Db --PPALREPLNNAASPGL--GAQPLPGHL-----PSHAMGGQMGGLPPG 345

QY 120 -QSPHY-PPHSHVLPASSAQSPQPIAPAPDRDRADFNGLPSCAFSYSDGKPGQWD 177
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 346 PEKGTPLAPSHS-LPSSS-----APAPMRFYSSSSSSSSSSSSSSSSAS 397
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 178 FVAANGAAP-YPGKDSRPTQV-VSQGRGILPSVPGRATPTVNGTCKNTTIPAKDA 235
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 398 FFPASQALPSYHSPFPPTTSLSVSNQPKVTPSLPSQAV-----WS 439
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 236 DGKFPENCKNTYLHAXHLRHLRHTGDRPYMCVLCCKDTFSRDLKRFQKCSIRGN 295
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 440 QGPPPPPPYGRLLANSN-----AHFGPPF-----P 464
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 296 PTGATHLSHP--NAHVKSQQQAANPVKPVQDEVSTVPPNGIPGTTYGBGAVNGN- 351
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 465 STGQASTAHFVPVTHHHHHQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 524
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 352 -----GLAPRPG-----YADHQTGMFPMSVNGMGRQGPEDAFPGGRPH 391
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 525 YAMSPSLGLSRYPYPPGPAHLPPHSHSVSYSAQAGPVGPPVSSSSSSSTSGQSYPCSHPS 584
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 392 -----QGAPRQAP-----KQSPYLVPQADPSCHQNLNDRNIEQVQPVQDPKR 437
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 585 PSQGPQAGYPPFPVPTVTSSATLSVIATVASSPAGY-----KT 625
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 438 PVMPGHGHPGEADWTSMFQAPQEGYMSQSPGQGEPIAHVETERKYPTTTAGQES 497
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 626 ASPGPPPYKRAPSPCAKATAPPY-----KEGSP-----PSFRGTTP 666
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 498 GNGGLYLASTMSGDGTQVP 516
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

QY 667 GYRG---TSPPAGEGTFPK 682
Db PTKPPTT--EVGGNLSAP-----PPANFHVTPNLP-----303

RESULT 5
PRDI_MOUSE STANDARD; PRT; 856 AA.
AC Q06636;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PR-domain zinc finger protein 1 (Beta-interferon gene positive-
GN regulatory domain 1 binding factor) (BLIMP-1).
GN PRDI OR BLIMP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA MEDLINE=94221646; PubMed=8168136;
RX Turner C.A., Mack D.H., Davis M.M.;
RT "Blimp-1, a novel zinc finger-containing protein that can drive the
RL maturation of B lymphocytes into immunoglobulin-secreting cells.";
RN Cell 77:297-306(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA MEDLINE=20572522; PubMed=11121475;
RX
```

Db 644 TQLAHLQXHYLVHTGCKHECQVCHKRFSTSNLKLHLHSGEKPYCKVCPAKFTQV 703
 QY 301 HLS-HPNAHVKSQQAA 317
 Db 704 HLKHLKRLHTRERPHKCA 721

RESULT 6

STEAM EMENI
 ID STEAM EMENI STANDARD; PRT; 692 AA.
 AC 074252;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor SteA.
 GN STEA OR STE12.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=162425;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC STRAIN=FGSC 4;
 RA Vallim M.A., Miller K.Y., Miller B.L.;
 RA "Aspergillus SteA (Sterile12-like) is a homeodomain-C2/H2-Zn-2 finger
 RT transcription factor required for sexual reproduction.";
 RL Mol. Microbiol. 36:290-301(2000).
 CC -!- FUNCTION: Transcription factor involved in sexual reproduction.
 CC Required for cleistothecial development and ascosporeogenesis. Not
 CC required for asexual reproduction (condidation). May act to
 CC repress medA expression.
 CC -!- PATHWAY: Involved in the mating pathway.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- DEVELOPMENTAL STAGE: Expression is low in vegetative hyphae and
 CC increases during conidiophore development.
 CC -!- SIMILARITY: BELONGS TO THE STE12 TRANSCRIPTION FACTOR FAMILY.
 CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF080600; AAC31206.1; -;
 DR HSSP; P07248; 1ABD
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0046564; P:negative regulation of transcription from P...; IMP.
 DR GO; GO:0000122; F:negative regulation of transcription from P...; IMP.
 DR GO; GO:0019953; P:sexual reproduction; IMP.
 DR InterPro; IPR003120; TF STE.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF02200; STE; 1.
 DR Pfam; PF00096; zf-C2H2; 2.
 DR ProDom; PD000003; ZnF_C2H2; 2.
 DR SMART; SM00424; STE; 1.
 DR SMART; SM00355; ZnF_C2H2; 2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
 KW Transcription regulation; DNA-binding; Repressor; Nuclear protein;
 KW Zinc; Metal-binding; Zinc-finger.
 FT DNA_BIND 56 165 BY SIMILARITY.
 FT ZN_FING 564 588 C2H2-TYPE.
 FT ZN_FING 594 616 C2H2-TYPE.
 SQ SEQUENCE 692 AA; 76694 MW; A0B86A6878437C13 CRC64;
 Query Match 7.9%; Score 225; DB 1; Length 692;
 Best Local Similarity 22.5%; Pred. No. 0.00012;
 Matches 115; Conservative 46; Mismatches 145; Indels 206; Gaps 26;

QY 4 RNHPRRPSTSLPQSSALPLSAPI-----SSMPMPQYTMQP-----QY 41
 Db 232 RTVDAMPPEQMAPO-MAP-PSIPLLTDESANSQMTYSIPMPNPPIQNLIKREADYGAIQY 289
 QY 42 -----PVSQPHLPLPQHHSQ-----SPAPHSYMGQ-----PPYRPD 74
 Db 290 DRNGMPIARIH-----QRHSSMPTVEYSPAPSVSSQYEDYSNRLSGLSPEVTPPPQSHH 344
 QY 75 LNRYPAA--SSHDVYASSAAPIMPHTTVGS-----LPPTSLFLSHFNPOAAQAQSPHY 125
 Db 345 IGPEPAYIANEDTGLYTAIPEM---STSGGFNPMVHMLPSNL-----ASAHP 388
 QY 126 PPP-----HS-VLPASSAQSPQIAPAPPRDRADFNGL-----PSGAFSYSD 170
 Db 389 PTPARTFHSNVYSVLEGSFTYKQ-----RRRRSSIPPGAANPVTTGTHSPAPALSAA 441
 QY 171 GKPGQWDPVAANGAAPYPKDS-----PRTQVVGSGQRRGI-----206
 Db 442 HKPSDLRRSVSSVAPGDTDSRHESLHRSVNSTYTATLPQKNLHMEGSRNGTFLSSVGE 501
 QY 207 -----LPSVPGRATPVTVNGVGTG-----KNVTI-----PAKD 234
 Db 502 HREQSSIPLTQAEIDLPAFANGTVESGAPNGVGHKSDRYATGVRARRASATMMELGPYPQ 561
 QY 235 ADGKEPCPCNKTYLHAKHLKRLHHTGDRPYCMVLCCKTDSRDLKXHFQKCSIRR- 293
 Db 562 KXHSCEPIFSCGRLEFKRLEHLKRLHRTHTQERPYPQPCNKAFSRSDNLAQH-----RKI 615
 QY 294 -----GNPTGATHLSH-----PNAHVKSQQAAANPVQDEVSSVTPPENGIFG 340
 Db 616 HEAQDQGP-----LVHEDDLNDDNESVSHDESPSESVHP-----AVPG 657
 QY 341 TTYGEGAVNGMLAPARFYADHQTGMFPMGS 372
 Db 658 V-----HGWTSMPSVALQSTMGSMWGS 680
 RESULT 7
 EGR2 MOUSE
 ID EGR2 MOUSE STANDARD; PRT; 470 AA.
 AC P08152;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Early growth response protein 2 (EGR-2) (Krox-20 protein).
 GN EGR2 OR EGR-2 OR KROX-20 OR ZFP-25.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RX MEDLINE=89219073; PubMed=2496302;
 RA Chavrier P., Janssen-Timmen U., Mattei M.-G., Zerial M., Bravo R.,
 RA Charnay P.;
 RT "Structure, chromosome location, and expression of the mouse zinc
 RT finger gene Krox-20: multiple gene products and coregulation with the
 RT proto-oncogene c-fos";
 RL Mol. Cell. Biol. 9:787-797(1989).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RX MEDLINE=8196089; PubMed=3129290;
 RA Chavrier P., Zerial M., Lemaire P., Almendral J., Bravo R.,
 RA Charnay P.;
 RT "A gene encoding a protein with zinc fingers is activated during
 RT G0/G1 transition in cultured cells.";
 RL EMO J. 7:29-35(1988).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RC STRAIN=FVB/N; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J.J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [4].
RP SEQUENCE OF 339-417 FROM N.A.
RX MEDLINE=8216608; PubMed=945975;
RA Chavrier P., Lemaire P., Revelant O., Bravo R., Charnay P.;
RT "Characterization of a mouse multigene family that encodes zinc
RT finger structures."
RL Mol. Cell. Biol. 8:1319-1326 (1988).
RN [5].
RP FUNCTION.
RX MEDLINE=90214631; PubMed=1969796;
RA Chavrier P., Vesque C., Galliot B., Vigneron M., Dolle P., Duboule D.,
RA Charnay P.;
RT "The segment-specific gene Krox-20 encodes a transcription factor
RT with binding sites in the promoter region of the Hox-1.4 gene."
RL EMBL J. 9:1209-1218 (1990).
RN [6].
RP FUNCTION.
RX MEDLINE=91234777; PubMed=1674431;
RA Gilardi P., Schneider-Maunoury S., Charnay P.;
RT "Krox-20: a candidate gene for the regulation of pattern formation in
RT the hindbrain."
RL Biochimie 73:85-91 (1991).
RN [7].
RP DOMAINS.
RX MEDLINE=92285130; PubMed=1598206;
RA Vesque C., Charnay P.;
RT "Mapping functional regions of the segment-specific transcription
RT factor Krox-20."
RL Nucleic Acids Res. 20:2485-2492 (1992).
CC -!- FUNCTION: SEQUENCE SPECIFIC DNA-BINDING TRANSCRIPTION FACTOR.
CC BINDS TO TWO SPECIFIC DNA SITES LOCATED IN THE PROMOTER REGION OF
CC HOX-1.4. MAY PLAY A ROLE IN THE REGULATION OF HINDBRIN
CC SEGMENTATION, MIGHT ACT IN COMBINATION WITH THE HOX NETWORK TO
CC SPECIFY ODD AND EVEN RHOMBOMERES, AND MIGHT PARTICIPATE IN THE
CC CONTROL OF THE EXPRESSION OF SOME OF THE HOMEOBOX CONTAINING
CC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P08152-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P08152-2; Sequence=VSP_006864;
CC -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN ADULT THYMUS AND EMBRYONIC
CC NERVOUS SYSTEM.
CC -!- INDUCTION: ACTIVATED DURING G0/G1 TRANSITION IN CULTURED CELLS.
CC -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24377; AAA393179.1; -.
DR EMBL; M24376; AAA393179.1; JOINED.
DR EMBL; M24377; AAA39380.1; -.
DR EMBL; M24376; AAA39380.1; JOINED.
DR EMBL; M06746; CAA29921.1; -.
DR EMBL; BC009093; AAH09093.1; -.
DR EMBL; M20759; AAA39381.1; -.
DR PIR; A30136; A30136.
DR PIR; S00256; S00256.
DR HSP; P08046; IAAI.
DR TRANSFAC; T00454; -.
DR MGD; MGI:95296; Egr2.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf C2H2; 2.
DR SMART; SMC0355; Znf C2H2; 3.
DR PROSITE; PS00028; ZINC FINGER C2H2_1; 3.
DR PROSITE; PS0157; ZINC FINGER C2H2_2; 3.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Repeat; Zinc-finger; Metal-binding; Alternative splicing.
FT DOMAIN 167 173 POLY-PRO.
FT ZN_FING 337 361 C2H2-TYPE 1.
FT ZN_FING 367 389 C2H2-TYPE 2.
FT ZN_FING 395 417 C2H2-TYPE 3.
FT VARSPPLIC 1 50 Missing (in isoform Short).
FT /FTId=VSP_006864.
SQ SEQUENCE 470 AA; 49819 MW; 67712733C77960F0 CRC64;
Query Match 7.8%; Score 223; DB 1; Length 470;
Best Local Similarity 22.0%; Pred. No. 0.0001;
Matches 95; Conservative 49; Mismatches 156; Indels 132; Gaps 17;
QY 9 RPSTSLPQGSAPLSPAPISS--MPMPQYTMQYPP-----HSYMGQPPYRDLNRYPASSHDVYA 87
DB 69 RLDELPPSSFAPL-SAPRNQTFYMGKFSIDPQPGASCYPEGIINIVSAGILQGVTPP 127
QY 47 HTLPPLQPHHSQSPAP-----HYPMPQYTMQYPP-----YSGCTGLYQ 183
DB 128 ASITASSSVTSASPNPLATGPGVCTMSQTPELDHLYSPPPPPP-----YSGCTGLYQ 183
QY 88 SSAAPIMPHITVGS-----LPTSFSLSHNPQQAQAQSPHY-----PPHSVLP 133
DB 184 DPAFLSPFSTTSTSLAYQPPSPSPKAMPDGLIPMIIDYPPGFPSPCQDRPHGAAG 243
QY 134 PASSAQSYPQPI-----APAPPRDRRADDNNGLPS-----GAFSYSDGKPGQWDVAA 181
DB 244 P--DRKFPFCLSLRVPPPLTSLIRNFTLGGPGAGVTGPGASGGGEGPRLPGSSAA 301
QY 182 NGAAPYPGKDSPTQVYQSGRRGILPSVFGRAPTVTNGVNGTGKNTTIPAKDADGKFC 241
DB 302 VTATYNPHPLPLRPIL---RPRKYENRPSK-TPV-----HERPYPC 339
QY 242 P--NCNKTYLHAKHLKRHLRLHTGDRPYMCVLCKDTFSRSDILKRHPQKSIIRGPTGA 299
DB 340 PAEGCDRRFRSDELTRHRIHTGKPFQCRICMRNFRSRSDLHTTHIR----- 387
QY 300 THLS-----HPNAHVKSQQAANPVKQVQDEVSVTPPPNGI 338
DB 388 THTGKFPACDYCGRKFAKSDERKRTKHLRQKERSA-PSAPFSAQSSASGPGSSQA 446
QY 339 PGTTYGEGAVNG 350
DB 447 GGSICGNSAIGG 458
RESULT 8
YQ36 CAEEL
ID YQ36 CAEEL
AC Q09457; STANDARD; PRT; 963 AA.

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative cuticle collagen C09G5.6.
 GN C09G5.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]_SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RA Palmer S.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like
 CC proteins. The cuticle functions both as an exoskeleton and as a
 CC barrier to protect the worm from its environment (By similarity).
 CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the
 CC cuticle by disulfide bonds and other types of covalent cross-links
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to the cuticular collagen family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z46791; CAA86755.1; -;
 DR PIR; T19140; T19140.
 DR WormPep; C09G5.6; CE01486.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR002486; Col_cuticle_N.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01484; Col_cuticle_N; 1.
 DR Pfam; PF01391; Collagen; 3.
 DR ProDom; PD000007; Clg_helix; 2.
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
 FT Multigene family; Collagen.
 FT DOMAIN 392 423 TRIPLE-HELICAL REGION.
 FT DOMAIN 441 503 TRIPLE-HELICAL REGION.
 FT DOMAIN 506 567 TRIPLE-HELICAL REGION.
 FT DOMAIN 663 666 POLY-PRO.
 FT DOMAIN 685 688 POLY-PRO.
 SQ SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;
 Query Match 7.7%; Score 222; DB 1; Length 963;
 Best Local Similarity 21.4%; Pred. No. 0.00022;
 Matches 145; Conservative 56; Mismatches 205; Indels 272; Gaps 36;
 QY 7 PSRPSTSLPQGS-----AFLPSAPIS-----SMPPQYTMQPOY----- 41
 Db 191 PQPSTGAPHSNNRTSLYNPQPPKTYGTNRVFNVPNPQNYTRQTYPDNRAPY 250
 QY 42 -PVSPHLPPLPQPHSQSPAPHSYNGQ-PPYRDL--NRYPASHDYVASSAIPMHT 97
 Db 251 KPTRSFNTPPPRQP-----SGYDSGQTPPSPSRINYNTRRNHGPYPEDQVPTAPPV 305
 QY 98 T-VGSLPTTSFLSHENPQAQAQSQSHYP-PPHS-----VLPPASSAQSYEQ----- 143
 Db 306 PQQRVPPTQTRNPNFTNTRQ-----PSRPVPTSDGHIEATTPNPSAQYPTGKRGSH 361
 QY 144 ---PIAPAP-----PRDRADFNGLPSG-----AFSYDSG----- 171
 Db 362 GFGQPRPRGTPRGNPCDCAQNHCPSPGPRGPRGPRGPRGPRGPRGPRGPRGPRG 421
 QY 172 ---KFGQWDV-----AANGAAPPKGD----- 191
 Db 422 SGVQSSYDPVIGCVQCPIGPPGERGPDGTGVPGEDGIDGEGVNGQDQGPAGDAPGY 481
 QY 192 -----SPRT-----QVGSQGRGLPSVGR-ATFVTGVNGTKNTTI 230

RESULT 9

SP7_HUMAN
 ID SP7_HUMAN STANDARD; PRT; 431 AA.
 AC Q8TDD2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transcription factor Sp7 (zinc finger protein osterix).
 GN SP7 OR OSX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]_SEQUENCE FROM N.A.
 RP SEQUENCE=Osteosarcoma;
 RA Ganss B.W.;
 RT "cDNA sequence, gene structure and chromosomal localization of the
 RT human osterix (OSX) gene."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]_SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Nakashima K., Zhou X., de Crombrughe B.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]_SEQUENCE FROM N.A.
 RP PubMed=14504442;
 RA Milona M.-A., Gough J.E., Edgar A.J.;
 RT "Expression of alternatively spliced isoforms of human Sp7 in
 RT osteoblast-like cells."
 RL BMC Genomics 4:43-43(2003).
 CC -!- FUNCTION: Transcriptional activator essential for osteoblast
 CC differentiation. Binds to Sp1 and E2F consensus sequences and to
 CC other G/C-rich sequences (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

Db 482 HGMGSPGTGPKGLPGRNGQSKIPGPPGQPGVM-GVPGRDGDPGTGEGHQDQSPGI 540
 QY 231 ---PAKD-----ADGKF-PCPNKNTY-----LHAKHLKRHLRLRHTGD 264
 Db 541 QQPGRDGTSGPDGQGVSAPEGAPGTGGYCFCKPKRSSKDFDNDAYNDDEKRGLEHR-P 599
 QY 265 RPYMCVLCKDFTSRSDILKRHFQKCSIRGNFTGATHLSHPNAHVRSQQOAAANPKPV 324
 Db 600 RGYDSERAEPRPQTVRTNTYDENS-----GAHQRRPNY----- 635
 QY 325 QDEVSTVPPNGIPGTTTGECAVNGGLAPARPGVADHDTMGTFPMSSVNGMGRGOPEDA 384
 Db 636 --EPSEAIVAPP-----QDYEDEVRREP-----PPKRP 663
 QY 385 PFGGRPHGAPWQAPKQSPYLQVGFADPSGHQLNDRNIEQV----- 427
 Db 664 PP---PHRQTPELYPEQPYVRRPPQNRGNVEVSREVPEAPRPGQVHSSGYG 720
 QY 428 -----KQPVQDPKRPVMPGHPGHEGELDWTSMFQ-----QAPE-GY-----MFSQS 469
 Db 721 GDDRKEQPKYME-SRPV--DEPRY--ETDAPSRPLKKKVIIRHPERGDRRQPSYEDS 775
 QY 470 MEGGOEPIHAHVETERY 487
 Db 776 KPRQEEPRRYETEAPRY 793


```

CC CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF477981; AAL84281.1; -.
DR EMBL; AF466179; AAC33377.1; -.
DR EMBL; AY150673; AAN85556.1; -.
DR Genew; HGNC:17321; SP7.
DR MIM; 606633; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat; Activator.
FT ZN_FING 294 318 C2H2-TYPE 1.
FT ZN_FING 324 348 C2H2-TYPE 2.
FT ZN_FING 354 376 C2H2-TYPE 3.
SQ SEQUENCE 431 AA; 44994 MW; 454A6FEA84309FF9 CRC64;

Query Match 7.7%; Score 221.5; DB 1; Length 431;
Best Local Similarity 24.1%; Pred. No. 0.00011;
Matches 111; Conservative 30; Mismatches 139; Indels 181; Gaps 22;

QY 11 PSTSL-----PQSAPLPSA-----PIS-SMPMPQYTMQPYVSQPHLPLQPHHS 57
DB 67 FFTSTNGLSPAGSPAPTSGYANDYPPFSGPGTGTQDPGLLVPKGHS-----S 118
QY 58 QSPAPHSYMGQPPYRPNLNPYPASSHDVYASSAAPIMPHITVGLSLPPTSLSHP----- 111
DB 119 SDCLPSVYTSILDMTHPYGSKYKAGIH-----AGISGPG--GNTPTPWDMHFGGNWLG 169
QY 112 NPAQAQAQ-----SPHYPPHSHVLPASSAQSYPQIAPAPRRDRADFNGLPSGAF 166
DB 170 GGQGGGGLGTLPTGPAQPLNPLP-----TYPSPDAFLNP----- 207
QY 167 SYSDEKFGQWDPVANCAAPYPG-----KDSPTQVVGSGQR---RGILPSV 210
DB 208 -----APYAPHLQPGQHVLPQDVYKPAVGNSSQLEGSGGAKPP 249
QY 211 PGRATPVTVNGVGTGKNTTIPAKDADGKFP-----P 242
DB 250 RGASTGSGGSGGSGA-----GRSSDCPCNCOELRGAAGAAKKPIHSHCHIP 299
QY 243 NCKNTYLHAKELKHLRLHRTGDRPYMC--VLCKDTFSRSDILKRFQ-----KCSIR 292
DB 300 GCGKVKYKASHLKAHLRWHTGERDFVCNWLFCGKRFTRSDERLHVHRTHTREKKFTCLLC 359
QY 293 RGNPTGATHLS-HENAHVKRSQQQAANPKYQVDEVSSTVPPNPGIPGTYGEGAVNGN 351
DB 360 SKRFTSRDHLKSHQRTGEB-----PGPGPPPSG--PKELGEGRSTGE 399
QY 352 GLAPARPGYADHQTGMFPMSVNGMRGQGPEDAPPGGRPHQ 392
DB 400 EASQTR-----PSAS-----PATFEKA-PGGSPEQ 425

RESULT 10
YANB SCHPO STANDARD; PRT; 582 AA.
ID YANB SCHPO STANDARD; PRT; 582 AA.
AC Q10076;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein C3H1.11 in chromosome 1.
CN SPAC3H1.11
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.

```

```

STRAIN=972;
MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadiou E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Nucleus (Potential).
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z68144; CAAG2264.1; -.
DR PIR; T38743; T38743.
DR HSP; P07248; 1ARD.
DR GeneDB SPombe; SPAC3H1.11; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 2.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
KW Hypothetical protein; Nuclear protein; Zinc-finger; Metal-binding;
KW DNA-binding; Repeat.
FT ZN_FING 423 446 C2H2-TYPE 1.
FT ZN_FING 452 473 C2H2-TYPE 2 (ATYPICAL).
SQ SEQUENCE 582 AA; 64125 MW; D325324E522E1B8D CRC64;

Query Match 7.7%; Score 221.5; DB 1; Length 582;
Best Local Similarity 22.2%; Pred. No. 0.00015;
Matches 107; Conservative 57; Mismatches 176; Indels 143; Gaps 23;

QY 18 GSAPLPSAP-ISSMPMPQYTMQPY-----YPVSQPHLPLQPHHSQSPAPH 63
DB 142 GFQPPFQNGISKSRIAQHQPSQTYDDTVDSFFDYWKAGAHNLAPQSSHTE--ASQ 199
QY 64 SYMGQPPYRPNLNPYPASSHDVY--ASSAAP-----IMPHIT-----TVGSLP- 103
DB 200 GYMYSTNTAHDATDIP-SSFNFYNTQASTAPNPQEIYQWSHEYRPHYQNNLLRAQPN 258
QY 104 -----PTSLSHP-----NPOAQQAQSSPHYPPHSHVLPFA-----SSAQSY 141
DB 259 VNCENFTTVPNYPPFQPSYNFNALV-----PSYTTLVSQLPSPCLTVSSGFLSTASSI 313
QY 142 PQ-----PIAPAPRRDRRRADFNGLPSGAF-----SYSDGKPGQ--WDP 178

```

```
Db 314 PSCNCSFVSKSGSPSYHABQEVNVSNGGIPSTSYNDTPQQSVTGSYNSGETMSTYINQ 373
QY 179 VAANGAAPYPKDSDPTQVVGSGRGILPSPVGRATPVTVNGVTGKNTTIPAKDADGK 238
Db 374 TNSGRSP---NSMEATEQITGITDGSKKRRR-----QPSNRKTSVPSRPGGKS 422
QY 239 FPCPNCKTYLHAKHLKRLLR-HTGDRPVMCLCKDTFSRDLKRRHFQKCSIRGNFT 297
Db 423 FVCEPSCKFKFKESEHLRHRSHLHTEKEFPVCI-CGRFRSRDNLKQH-----ERLHYN 475
QY 298 GATHLS---HPNAHVRSQQQAANVPKPVQDEVSTVPPNGIPGT----- 341
Db 476 ASPRLACFPQSGYYSSGAPGAPVQKPIED--LNKIPINQMGDSQIENIMLSSOR 533
QY 342 -----TYGGAUNGLAPARGYADHQTGMGFMSSVNGMGKQGPEDAPGGRPHOGAPW 396
Db 534 PLSQQTVPBIAAPNSIRPELLSKLPVQPNOKMPLNPM-----HOYQPY 579
QY 397 POA 399
Db 580 PSS 582

RESULT 11
EGR2 RAT
ID EGR2 RAT STANDARD; PRT; 470 AA.
AC P51774;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Early growth response protein 2 (EGR-2) (Krox-20 protein).
GN EGR2 OR EGR-2 OR KROX-20 OR KROX20.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Zaharic I., Durtschi B.A., Mason-Parker S.E., Abraham W.C.,
RA Tate W.P.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=96192065; PubMed=8619872;
RA Inokuchi K., Murayama A., Ozawa F.;
RT "mRNA differential display reveals Krox-20 as a neural plasticity-
regulated gene in the rat hippocampus.";
RL Biochem. Biophys. Res. Commun. 221:430-436(1996).
CC -!- FUNCTION: SEQUENCE SPECIFIC DNA-BINDING TRANSCRIPTION FACTOR.
CC BINDS TO TWO SPECIFIC DNA SITES LOCATED IN THE PROMOTER REGION OF
CC HOX-1.4 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC isoId=P51774-1; Sequence=Displayed;
CC Name=Short;
CC isoId=P51774-2; Sequence=VSP 006865;
CC -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U78102; AAB36783.1; --.
EX MEDLINE=97388474; PubMed=9247308;
```

```
DR EMBL; D83508; BAAL1932.1; -.
DR PIR; JC4716; JC4716.
DR HSSP; P08046; 1AAY.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Repeat; Zinc-finger; Metal-binding; Alternative splicing.
FT DOMAIN 157 173 POLY-PRO.
FT ZN_FING 337 361 C2H2-TYPE 1.
FT ZN_FING 367 389 C2H2-TYPE 2.
FT ZN_FING 395 417 C2H2-TYPE 3.
FT VARSPLIC 1 50 Missing (in isoform Short).
FT FTID=VSP 006865.
SQ SEQUENCE 470 AA; 49849 MW; 2B28185752A91F76 CRC64;

Query Match 7.7%; Score 221; DB 1; Length 470;
Best Local Similarity 22.5%; Pred. No. 0.00013;
Matches 97; Conservative 48; Mismatches 155; Indels 132; Gaps 18;

QY 9 RPESTSLPQGSAPLPSAPISS--MPMPQYTMQOYP-----VSQP 46
Db 69 RLDLPYESSFAPL-SAPRNQFTTNGKFSIDPQYEGASCYPEGINIVSAGILQGVTPP 127
QY 47 HTLPQLQPHHSQSPAP-----HSYMGQPPYRFDLNRYPASSHDVYA 87
Db 128 ASTTASSSVTSASPNPLATGLVCTMSQTQPELDHLYSPPPPPP-----YSGCTGDLQ 183
QY 88 SSAAPIM--PHTTVGSL---PPTSFLSHPNQQAQAQSQSPHY-----PHSVLP 133
Db 184 DSAFLSPPTTSTSSLAYPPSPKPMADPGLIPMDYPGFFPSPQCQDHPHGAAG 243
QY 134 PASSAQSPQPI-----APAPPRDRADENGLPS-----GAFSYSQKQGWDPVAA 181
Db 244 P--DRKPPFCLDSLRLVPPPLTSLTRNFTLGGESAGVTGSGSGGEGPLPSGSGAA 301
QY 182 NGAAPYPGKDSPTQVVGSGRRGILPSPVGRATPVTVNGVTGKNTTIPAKDADGKFP 241
Db 302 VTATPYNPHLPLRPIL---RPRKYFNRPSK-TPV-----HERPYPC 339
QY 242 P--NCNKTYLHAKHLKRLLRHTGDRPYMCVLCCKDTFSRDLKRRHFQKCSIRGNPTGA 299
Db 340 PAEGCDRFRSDELTRHRIHTGHKPPQCRICMNFNFRSDHLATTHIR----- 387
QY 300 THLS-----HPNAHVRSQQQAANVPKPVQDEVSTVPPNGI 338
Db 388 THTGEKPFACDYCGRKFAFSRDERKRRHTKHLRQERKSSA-FSSASASQSSASGPGGSOA 446
QY 339 PGTTYGEGAVNG 350
Db 447 GGSLCGNSAIGG 458

RESULT 12
MLL2 HUMAN
ID MLL2 HUMAN STANDARD; PRT; 5262 AA.
AC O14686; O14687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related
DE protein).
GN MLL2 OR ALL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
EX MEDLINE=97388474; PubMed=9247308;
```



```

FT FT VARSLIC 1454 1454 /FTID=VSP_008559.
FT FT TISSUE=Cerebellum, and Striatum; E -> EGET (in isoform 3).
FT FT MEDLINE=97317138; PubMed=9173996; /FTID=VSP_008560.
FT FT VARIANT 4949 4949 R -> H (in dbSNP:3782356). /FTID=VAR_017115.
FT FT SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417B44 CRC64; /FTID=VAR_008559.

Query Match 7.6%; Score 218.5; DB 1; Length 5262;
Best Local Similarity 22.1%; Pred. No. 0.0017;
Matches 146; Conservative 58; Mismatches 229; Indels 229; Gaps 32;

QY 3 PRNHP-----SRPSTSLPOGSAP-----LPSAPISMPM-PQYTMQ 39
DB 2087 PSHDPIFRPSYTDYACPLTPRPQPPPPSCCALPPRLSPSFFRVPVSPQSS 2146
QY 40 QPVSVQHTL-----PPLQPHSQSPAPHSYMQPPYRPLNRYPASSHDVYASSAP 92
DB 2147 QSPLT-PRPLSAEACPSVTPRF-QSDPDYS--RPFSRP-----QSRDPFAPLHKP 2194
QY 93 IMP-----HTVVG-----SLP-----PTSLSHEN----- 112
DB 2195 PRQPPEVAFKAGSLAHTSLGAGGFPALPAGFAGELHAKVPSGQPPNFVRSPTGAFVG 2254
QY 113 -----PQAAQAQSPHPPPHSLVLPASSAQSYPOPTAPAPPRDRRADFN----- 158
DB 2255 TFSMRFTFPQAVGFSLKLP--PVPQGLPPPHGNHSH---FGPQTLGKPOSTNYTAT 2309
QY 159 -----NGLPSGAFSYSDGKPGQGDWVAANGAAPPYPKDSPTQVY-GSQGRGILPSVPG 212
DB 2310 GNHFSGSGLPSSGSGTGESYGLSLRPPSVLPFPAPDGSLLPYLSHGASQSGITSPVEK 2369
QY 213 RATPVTVNGVNGKNTTIPAKDAG-KFP-CPNCKNTYLHAKHLRHLRHHTGDRPYMCV 270
DB 2370 REDP-----GTGMSLATAELPGTQDPGMSGLSQTELEKQORQL----- 2411
QY 271 LCKDTFSRDLKRFKCSIRGNPT-----CATHLSHPNAHVKSQQA 316
DB 2412 -----RELLRQIQIORTLRQKETAATAAAGAVGPPGSGWAEPSPAFEQLSRGQ--- 2461
QY 317 AANPKVPQDESVSTVPNGIPGTYGEGAVNGNL-----APAPGYAD---HQTMG- 367
DB 2462 --TPAGTQDKSLVGLPFSKLSGLPGSPSDDLRLSRPPFPATPSMDVNSQLVGG 2519
QY 368 -----FPMSSVNGMGR-----GQP- 381
DB 2520 SQAFYQRAYPGSLPLQOQQOQQOQQOQATAATSMRFAMSAFPSTPGPELGRQALGSPL 2579
QY 382 ---EDAFGGRPHQAPWPAQKQFYL---VOPGADPSGHQNLNDRLEQVK-QPVVQD 434
DB 2580 AGISTRPLG--PGEVPVGGAGAQITELHNHVKGLGPGGTPFGCGPPQRPFPVSED 2637
QY 435 PKRPVMPGHPGPGELDWTSMFPQAPGEGYMFSGMPGQGEPIAHVETERYKYYTTTAG 494
DB 2638 PHRLAPEGLRGLA-----VSGLPQKP-----SAPPAPELNLSLHTPHTKGTLPTG 2685
QY 495 QE 496
DB 2686 LE 2687

RESULT 13
DRPL RAT
ID DRPL RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.,
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RT homologue."
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Eppien J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat."
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -! SUBUNIT: Interacts with WPL1 and WBP2 (By similarity).
CC -! TISSUE SPECIFICITY: Predominant neuronal expression, although
CC markedly reduced amounts are found in most other tissues.
CC -! DEVELOPMENTAL STAGE: Similar expression at all development stages
CC (14.5 dpc, 17.5 dpc, newborns and adults).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31777; AAA80337.1; -
DR EMBL; X89453; CAA61623.1; -
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT DOMAIN 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 589 589 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Query Match 7.4%; Score 212.5; DB 1; Length 1183;
Best Local Similarity 22.4%; Pred. No. 0.0008;
Matches 139; Conservative 44; Mismatches 22; Indels 217; Gaps 30;

QY 7 PSRPSTSLPO-----GSAPLPSAPISMPMPQYTMQPOYVPSQ-----PHTLPLPQPHS 57
DB 166 PSHDPIFRPSYTDYACPLTPRPQPPPPSCCALPPRLSPSFFRVPVSPQSS 220
QY 58 QSPAPHSYMGQPP-----YRPDLNRYPDASHDVYASSAPIM-PH 96
DB 221 PGSAGGVLSGPPMPKPGKGAASVGPSSGKGQHPPTTPIPISSG--ASGAPPAKPPN 278
QY 97 TTVG-----SLPPTSFLSHNP-----QAQO-----AQOSP-HYPPPHSV-- 131
DB 279 TPVGAGNLPSAPPATFPHVTFNLPPPPPALRLNNAASPPFGMAQPIGHLFSPHAMGQ 338
QY 132 -----LPPASSAQSYPOPTAPAPPRDRRADFNGLPSGAFSYSD 170
DB 339 GMSGLPQGEKGTLPASPFLPPASS-----AFGFP--MRYPFSCSSSSVAASSS 389

```

QY 171 GKPGQWDPVAANGAAMPYCKDQSPRTQV-VGSGRGGILPSVGRATFVINGVNGTGKNTT 229
Db GO:0005515; F:protein binding; IPI.
DR GO:0030528; E:transcription regulator activity; IDA.
DR GO:0045941; P:positive regulation of transcription; IDA.
QY 230 IPAKDADGKPCPCNCKTYLHAKLKRHLRHTGDRPYMCVLCKDTFSRDLKRHFQKC 289
Db IPI:000096; ZF-C2H2; 3.
441 PPPPPYGL-LPNN-THPGPP-462
QY 290 STRGNPTGATHLSHNA---HVRSQQAAANPKVQDEVSSVTPPPNGIPGTYGEG 346
Db PTGGQSTAHPPAPAHHHQOQQPQP-QPQQHHGNSGPPPG---AYPH 511
347 AVNGN---GLAPARG-----YADHQTMGFPMSSVNGMGROQ 380
512 LESSNHHAPYVMSLSGLRYPGPAPHLPSHGQVSYQAGNPGPVSSSSNSGSS 571
QY 381 PEDAFEGGRPH---QGAPWFOAPKQSPYLVPQGDPSGHQNLNDRNIEQVKQPVVQDP 435
Db SQAAYSCSHSPSSQPGQASYPFP---VPPITSSA---TLSTVIATVASSP 618
QY 436 KRPVMPCHPGHGLDWTSMFQDQAEQYMFSSQMGQEPHAEVETKRYPTTAGQ 495
Db AGYKTASPPGPP-----QYKRAEPPSGYKATATPPGYKQSP-----PSRTGT 662
QY 496 ESCMNGLYLASTWSGDGTQVP 516
Db PPGYRG---TSPPAGPGTKP 680

RESULT 14

KLF2_MOUSE
ID KLF2_MOUSE STANDARD; PRT; 354 AA.
AC Q0843;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kruppel-like factor 2 (Lung kruppel-like factor).
GN KLF2 OR LKLF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96025976; PubMed=7565748;
RA Anderson K.P., Kern C.B., Crable S.C., Lingrel J.B.;
RT "Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Kruppel-like factor: identification of a new multigene family."
RL Mol. Cell. Biol. 15:5957-5965(1995).
CC -!- FUNCTION: Binds to the CACCC box in the beta-globin gene promoter and activates transcription.
CC -!- SUBUNIT: Interacts with WPI1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN THE LUNGS AND SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; U25096; AAA86728.1; -.
DR HSSP; P08047; 1SP2.
DR TRANSPAC; T01677; -.

DR MGI:1342772; Klf2.
DR GO:0005515; F:protein binding; IPI.
DR GO:0030528; E:transcription regulator activity; IDA.
DR GO:0045941; P:positive regulation of transcription; IDA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; ZF-C2H2; 3.
DR ProDom; PD00003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 60 70 POLY-PRO.
FT DOMAIN 165 169 POLY-PRO.
FT DOMAIN 224 230 POLY-ALA.
FT ZN_FING 271 295 C2H2-TYPE 1.
FT ZN_FING 301 325 C2H2-TYPE 2.
FT ZN_FING 331 353 C2H2-TYPE 3.
SQ SEQUENCE 354 AA; 37700 MW; C4A99D018AC5BAF7 CRC64;
Query Match 7.4%; Score 211.5; DB 1; Length 354;
Best Local Similarity 26.4%; Pred. No. 0.00028;
Matches 92; Conservative 27; Mismatches 133; Indels 97; Gaps 17;
QY 5 NHPSPPSTSLPQGSAPLPSAPISSMMPQYTMQPYVPSQPHHTLPLOPHHSQSPAPHS 64
Db NPPEPPPPPPPAFYVPEPGAP-----PPYSI-PADSLGTELLRPDLDP--PQGPALRG 109
QY 65 -YMGQPPYRPDLNRYPASSHDVYASSAAPIMHTTVGSLPPTSFLSHNPQAAQAAQSQSP 123
Db RFLAPPGR--LVKAEPPVGGGVCAPGLAH-----GPRGLKLEGAPGATGCMRGP 161
QY 124 --HYPPPHSVLP-----PASSAQS-YPOFIAPAPPRDRRDNFNLPSGAFSYSDG 171
Db AGRPPTPTPLSPDGLRIPASGRNPPFPFGPGP-----SFGGPGALHYGPP 213
QY 172 KPQGWG-----PVAANGAAYPKDSPTQVSGQGRGILPSVPGRATPTVNGVNGT 224
Db AFGAFLFEDAAAAAALGLAP-----PATRGLTTPFSSPLE-- 250
QY 225 GKNTTIPAKDADGKPCP-----NCKNTYLHAKLKRHLRHTGDRPYMCVL- 271
Db -----LLEAKPKGRSRWFKRAATHCTCSYVNGKITKSHLKAHLRHTGKPYHCNWE 306
QY 272 -CKDTFSRSDILKRHFQKCSIRGNPTGATHL-----SHPNAHVR 311
Db GCGWKPFARSDLTTRHYRK---HTGHRPFOCHLCDRAFSRSDHLALHMKR 352

RESULT 15

PRDI_HUMAN
ID PRDI_HUMAN STANDARD; PRT; 789 AA.
AC O75626;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PR-domain zinc finger protein 1 (Beta-interferon gene positive-regulatory domain I binding factor) (BLIMP-1) (Positive regulatory domain I-binding factor 1) (PRDI-binding factor-1) (PRDI-BF1).
DE domain I-binding factor 1) (PRDI-binding factor-1) (PRDI-BF1).
GN PRDM1 OR BLIMP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91224495; PubMed=1851123;
RA Keller A.D., Maniatis T.;
RT "Identification and characterization of a novel repressor of beta-interferon gene expression."
RL Genes Dev. 5:868-879(1991).
CC -!- FUNCTION: Transcriptional repressor that binds specifically to the PRDI element in the promoter of the beta-interferon gene. Drives

```
CC the maturation of B lymphocytes into Ig secreting cells.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF084199; AAC33300.1; -.
DR PIR; A39564; A39564.
DR HSSP; P08048; 7ZNF.
DR Genew; HGNC:9346; PRDM1.
DR MIM; 603423; -.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR GO; GO:0000122; P:negative regulation of transcription from P. .; TAS.
DR InterPro; IPR001214; SET.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00856; SET; 1.
DR Pfam; PF00096; zf-C2H2; 5.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00317; SET; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
DR Transcription regulation; Repressor; DNA-binding; Zinc-finger;
KW Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 49 169
FT ZN_FING 539 561
FT ZN_FING 567 589
FT ZN_FING 595 617
FT ZN_FING 623 645
FT ZN_FING 623 645
SQ SEQUENCE 789 AA; 87991 MW; 6A1FEBEAD309BC7D CRC64;

Query Match 7.4%; Score 211; DB 1; Length 789;
Best Local Similarity 20.6%; Pred. No. 0.00065;
Matches 90; Conservative 44; Mismatches 152; Indels 150; Gaps 16;

QY 7 PSRPSTSLPGSAPLPSAPTSSMPPOYTWQPYPVSQP-HTLPPLQPHSQSPAPHSY 65
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 285 PTYITRSPISPTTSPARS-----PQSLKSSSPHSPTGNTVSPVPGSQEHRDSYAY 340
QY 66 MGQPPYRPLNRYPASSHVDYASSAAPIMPTTVGSLPPTFLSHPNPQAQAQSPHY 125
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 341 LNASYGTGLGYPG-----YAP-----LPH-----LPPAFTPSY-----NAHY 374
QY 126 PP-----PHSVLPASSAQS 141
DQ 375 PKFLLPYGMNCGLSAVSNMNGINNFGLPRLCPVYNNLGGSLPHPLNPTSLPSSL 434
QY 142 PQP-----IAPAPRRDRADFNNGLPSGAFSYSDGKPGQWDPVAA-NGAAYPGKDSPT 195
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 435 PSDGNARLLQEPHEPREVLV-----APHSAFSTGAASMKKAKCSPTSGSTAGTAATAE 490
QY 196 QVVGQGRGIL--PSVPGRATPTVNGVTG-KNTTIPAKDADGK----- 238
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 491 HVVQPKATSAAMAAPSSDEAMNLIKNNMTGYKTLPYPLKKQNGKIKYECNVCATFGQ 550
QY 239 -----FPCPNCKYTLHAKHLKRLHLLRHTGDRPYMCVLCKDTFSRSDIL 282
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 551 LSNLKVHLRVHSGERPFKQCTCKNGFTQLAHLQKHYLVHTGEKPCQVCHKRFSTSNL 610
QY 283 KHFFQ-----KCSIRGNPTGATHLS-HENAHVKESQQQ----- 315
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 611 KTHLRLHSGEKPYQCKQVFAKFTQVHLKHLKRLHRLRERPHKCSQCHKNYIHLCSLVHL 670
QY 316 -----AAANPVKPVQD 326
```

Db 671 KGNCAAAPAGLPLED 686

Search completed: April 1, 2004, 17:47:21
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:40:49 ; Search time 45 Seconds

(without alignments)
3638.977 Million cell updates/sec

Title: US-10-029-180-8

Perfect score: 2866

Sequence: 1 MDPNHPSPSTSLPQSGA.....NGLYLASTMGDTGTVQPARQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2866	100.0	519	3 Q8J177	Q8J177 emericella
2	274.5	9.6	1724	5 P91019	P91019 caenorhabdi
3	252.5	8.9	722	4 Q8IYCI	Q8IYCI homo sapien
4	240	8.4	723	4 Q8NBB9	Q8NBB9 homo sapien
5	240	8.4	1766	4 Q8NF45	Q8NF45 homo sapien
6	238.5	8.3	1191	4 Q86V38	Q86V38 homo sapien
7	237	8.3	1193	5 Q9VQ94	Q9VQ94 drosophila
8	233	8.1	722	3 Q9C170	Q9C170 neurospora
9	232.5	8.1	3084	12 Q8U211	Q8U211 pseudorabie
10	231.5	8.1	726	11 Q924S6	Q924S6 mus musculu
11	230.5	8.0	596	11 Q9CVF3	Q9CVF3 mus musculu
12	229.5	8.0	926	5 Q9W3G1	Q9W3G1 drosophila
13	224	7.8	839	16 Q9RX57	Q9RX57 deinoecocuc
14	222.5	7.8	1175	11 F70200	F70200 mus musculu
15	221.5	7.7	413	4 Q7Z718	Q7Z718 homo sapien
16	221	7.7	412	11 Q9QYG4	Q9QYG4 rattus norv

17	219.5	7.7	695	3 Q8NKD8	Q8Nkd8 gibberella
18	216.5	7.6	1778	4 Q8IWR5	Q8iwr5 homo sapien
19	216.5	7.6	1858	4 Q8NFE8	Q8nff8 homo sapien
20	216.5	7.6	1858	4 Q8IZD2	Q8izd2 homo sapien
21	216	7.5	389	4 Q86VQ7	Q86vq7 homo sapien
22	216	7.5	398	13 Q802P5	Q802p5 falco colum
23	215	7.5	428	11 Q811U1	Q811u1 rattus norv
24	214.5	7.5	1175	11 Q35126	Q35126 mus musculu
25	213	7.4	389	4 Q8N805	Q8n805 homo sapien
26	213	7.4	1966	5 Q8NHX6	Q8nhx6 drosophila
27	212.5	7.4	2133	5 Q8IQY4	Q8iqt4 drosophila
28	212	7.4	382	4 Q00599	Q00599 homo sapien
29	211.5	7.4	697	3 Q8NK75	Q8nk75 glomerella
30	211	7.4	417	13 Q90XX6	Q90xx6 oncorhynch
31	211	7.4	691	4 Q86WM7	Q86wm7 homo sapien
32	211	7.4	728	4 Q7Z6T5	Q7z6t5 homo sapien
33	211	7.4	997	5 Q9GYL4	Q9gy14 caenorhabdi
34	211	7.4	1153	5 Q8MQW5	Q8mgw5 drosophila
35	211	7.4	1966	5 Q8IOA6	Q8iga6 drosophila
36	211	7.4	1985	5 Q9VSK5	Q9vsk5 drosophila
37	211	7.4	1985	5 Q8T9N4	Q8t9n4 drosophila
38	211	7.4	1988	5 Q86BH2	Q86bh2 drosophila
39	210.5	7.3	745	16 Q89X06	Q89x06 brachyhirzob
40	210	7.3	389	13 Q802J6	Q802j6 deconychura
41	210	7.3	1350	13 Q91929	Q91929 xenopus lae
42	209.5	7.3	903	16 Q82HF3	Q82hf3 streptomyce
43	209	7.3	378	13 Q802I3	Q802i3 vermivora v
44	208.5	7.3	533	11 Q8CAT6	Q8cat6 mus musculu
45	208.5	7.3	838	5 Q8T5L5	Q8t5l5 anopheles g

ALIGNMENTS

RESULT 1

Q8J177 ID Q8J177 PRELIMINARY; PRT; 519 AA.
AC Q8J177;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE RfEC.
GN RfEC.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RA Holtzman D.A.;
RT "Enhanced Production of Secondary Metabolites in Filamentous Fungi by
RT Ectopic Expression of Regulatory Proteins";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF467711; AAO14631.1; -
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
SQ SEQUENCE 519 AA; 55648 MW; 4AACD0892BA3FA40 CRC64;

Query Match 100.0%; Score 2866; DB 3; Length 519;
Best Local Similarity 100.0%; Pred. No. 3.4e-196;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPNHPSPSTSLPQSGAPLPSAPISMPMPQYTMQPYVQSPHTLPPLQPHHSSP 60

Db 1 MDPNHPSPSTSLPQSGAPLPSAPISMPMPQYTMQPYVQSPHTLPPLQPHHSSP 60

QY 61 APHSYMQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLPSTSLSHENPQAQAQ 120

Db 61 APHSYMQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLPSTSLSHENPQAQAQ 120

QY 121 QSPHYPPHSVLPASSAQSYPQIAPAPPRDRADFNGLPSGAFSISDGKFGQWDPVA 180
 DB 121 QSPHYPPHSVLPASSAQSYPQIAPAPPRDRADFNGLPSGAFSISDGKFGQWDPVA 180
 QY 181 ANGAAPYKGDSPRTQVGSQGRGILPSVPGRATPVNGVNGTGNKTTIPAKDADGKEP 240
 DB 181 ANGAAPYKGDSPRTQVGSQGRGILPSVPGRATPVNGVNGTGNKTTIPAKDADGKEP 240
 QY 241 CPNCNKTYLHAKHLKHLRHTGDRPYMCLCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
 DB 241 CPNCNKTYLHAKHLKHLRHTGDRPYMCLCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
 QY 301 HLSHPNAHVRSQQAAPVQVDEVSITVPPNGIPGTTTGEAVNGNGLAPAPGY 360
 DB 301 HLSHPNAHVRSQQAAPVQVDEVSITVPPNGIPGTTTGEAVNGNGLAPAPGY 360
 QY 361 ADHOTMGFMSSVNGMGRGQPEDAFGGRPHQAGAPWQAPKQSPYLIVQPGADPSGHQJNI 420
 DB 361 ADHOTMGFMSSVNGMGRGQPEDAFGGRPHQAGAPWQAPKQSPYLIVQPGADPSGHQJNI 420
 QY 421 DRNIEQVKQVVDKPRVMPGHPGHPGELDTSMFQPAPEGYMSQSPGQGPPIHAH 480
 DB 421 DRNIEQVKQVVDKPRVMPGHPGHPGELDTSMFQPAPEGYMSQSPGQGPPIHAH 480
 QY 481 VETERKYPTTTAGQSGMGLVLA STMDSGDGTVPARQ 519
 DB 481 VETERKYPTTTAGQSGMGLVLA STMDSGDGTVPARQ 519

RESULT 2

P91019 ID P91019 PRELIMINARY; PRT; 1724 AA.
 AC P91019;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C01G8.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "The sequence of C. elegans cosmid C01G8.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80439; AB37646.3; -.
 DR PIR; T29266; T29266.
 DR WormPep; C01G8.9a; CE30415.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR001606; ARID.
 DR InterPro; IPR006031; XTPPX.
 DR Pfam; PF01388; ARID; 1.
 DR Pfam; PF02162; XTPPX; 7.
 DR SMART; SM00501; BRIGHT; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1724 AA; A981E4CEE388EFB8 CRC64;

Query Match 9.6%; Score 274.5; DB 5; Length 1724;
 Best Local Similarity 22.6%; Pred. No. 7.le-11;
 Matches 156; Conservative 53; Mismatches 188; Indels 293; Gaps 36;
 QY 16 POGSAPLPSAPITSSMPMPQYTMQ-----POXPVSQP-----HTLPPLQ-----PH 55
 DB 24 PKDDQVPPAPDASMTPTQOQQOQNLPPQPFYHPAQOHPHQHSVFNPSFAPPGQH 83
 QY 56 HSQSPA-----PHSY-MGQPPYRDLNRYPASSHDVYA----- 87
 DB 84 HFQHPGMPMEWRPPGAEYQMPGYPAGYPPYGMPPRHPHPHAYGYPGAPYGP 143
 QY 88 -SSAIPIMHTTVGSLPTPSFLSHNP---QAQAQAQSPHYPPH-----SVLPASSA 138
 DB 144 QMRRPMMAPAGDMVRMP-----GPTFTWAQQAQAASRVPFKEGNGNPAATSSS 198
 QY 139 QSYVPOPIAPAPPRDRADFNGLPSGAFSISDGKPGQMDP----- 178
 DB 199 QIPSPASSIAIESLDD---KPSGTMPAPPPPOQHPPPPQOIMSPMPQAAPSQ 254
 QY 179 -VAANGAAPPYKQDSPRT-----QVVG-----SQRRG----- 205
 DB 255 ATPSSAAASVAAPDTPKVSGLSKAELFEKLVGPTTLHNMKVMWAERGFERLIEFCE 314
 QY 206 -----ILPSYVGRATFTN---GVNGTKNTTIPAKDADGKFCPCNKT----- 247
 DB 315 HNGEPLTMVPSQKSIDLRLHYIGVRAKGGQOV--TKDKYWNKLTENPDLAESSAAG 373
 QY 248 YLHAKHLKHL-----RTGDRPYMCLCKDTFSRSDILKRHFQKCSIRRGNTGATHLS 303
 DB 374 YQLRKHVQRLHMLECRETGRN-----EDEVAFAKMKRQ-----RKREPAAG-- 417
 QY 304 HPNHVRSQQAAPVQVQ--DEVSTVPPNGIPGTTTGEAVNGNGLAPAPGYA 361
 DB 418 --NA-----AAAAAQAQGDQKQ--HGAPG-----GSGAPPPG-- 450
 QY 362 DHQWTFMSSVNGMGRGQPEDAFGGRPHQAGAPW--PQAPKQSPYLIVQPGADPSGHQNLN 419
 DB 451 -----PGAYPONGAPGPGPGPPGPPF--GHPGMDPNVHYQ 484
 QY 420 IDRNIEQVKQVVDKPRVMPGHPGHPGELDTSMFQFQ-----APEGYMSQSPG-- 473
 DB 485 -----QHGMMPPHPGHPG-----YPPQHMNAFSPGYPGHPGPPGP 521
 QY 474 -----QEPHAHV-----ETERKY----- 488
 DB 522 GGPFGPGQAMRAPMQHMQEMDSHQRYAHQAQAAAHHAQQAQAQAQAQAQAQASTP 581
 QY 489 -PTTTAGQSGMGLVLA STMDSGDGTVPQA 517
 DB 582 APSTTAPQAASS--QAAPTSGANT-QPA 608

RESULT 3

Q81YCI ID Q81YCI PRELIMINARY; PRT; 722 AA.
 AC Q81YCI;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to zinc finger protein.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC036105; AAH36105.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO:0004969; P-histamine receptor activity; IEA.
DR GO:0007196; P-G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR003980; H3_receptor.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; Zf-C2H2; 8.
DR PRINTS; PR01471; HISTAMINEH3R.
DR SMART; SM00355; ZnF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
SQ SEQUENCE 722 AA; 76881 MW; 8CA879DAB9F78759 CRC64;

Query Match 8.8%; Score 252.5; DB 4; Length 722;
Best Local Similarity 24.3%; Pred. No. 9.2e-10;
Matches 146; Conservative 45; Mismatches 152; Indels 257; Gaps 36;

Qy 7 PSRPSLPGSAPLPSAPLSSMPQYTMQPVYSQPHLPLQPHHSQSP-----A 61
Db 215 PERPLAAT---SAAPPP-----QFQPPPPQPE-PRSVQPEPEPEREATPTTAPAA 264
Qy 62 PHSYMGOPVYRPD-----LNRYPAS-----SH 83
Db 265 PE-----PPAPPEPRCVCQGSFQTSWFLKGMKHKASFDHACPVCGRCFKEFWFLKNH 320
Qy 84 -DVYASSAAPIMPTTGVSLPPTSLFSLHPNP-QAQAQAQSPH-----YPPHVSVLPPAS 136
Db 321 MKVEASKLGP-----LRAPGASGPAPAPQPPDLGLLAYEP-----LGPAL 361
Qy 137 SAQSYPOPIAP-PRRRADFNGLPSGAFSY-----SDGKPGQWDVVAANGAAPYCGK 190
Db 362 L-----LAPAPTAEKRE-----PFLGLYLRLRAGEGRNG-----EGAPPGPGR 402
Qy 191 D-----SPRTQVVGSGR-----GILPSV-----PGR 213
Db 403 SFGFRPLSSLPARARHRAEBEVEEVEABEETWARGRSLGSLASLAPRGEGPCH 462
Qy 214 -----ATPVING--VNGT-----GKNTTIPAKDADGKFPNCNKTYLHAKELK 255
Db 463 SASAAGAAQARSTATQENGLLVGTRPEGGRGAT--GKD-----CPFCGKSFRSAHELK 514
Qy 256 RHLRHRTGDRPVMVLCVDTFSRDLKRLHFKQKCSIRGNTGTHLSHPNAHVRS--- 312
Db 515 VHLRVHTGERPKYKCPHCDYAGTQSGSLKYHLQRHREQRSGAGFGPPPPPPSQGSAP 574
Qy 313 QQAQAANPVKQVDEVSSTVPPNGIPGTTTVEGAVNGVGLAPRGVADHCTGFWPMS 372
Db 575 QSGAKPSQAPATWEGASSPPPS-----SGAGPSRRKPASPG-----RTLR----- 617
Qy 373 VNGMGRGQPE-----DAPPGGRPHQ-----GAPW----- 396
Db 618 -NGRG-GEAEPLDLSLRAGPGGAGFGGALHRLCLFPCFATGAPELMALHLQVHHSRRARG 675
Qy 397 ---FOAPKQSYLVQPGADPSGHQNLNDRNEIQVQPVVQPKRPVMPCHPG-----HPGE 449
Db 676 RRPQADASPPY---ARVPSG-----ETPPPSQREGEGSLSRPGE 714

RESULT 4

Q8NBB9 PRELIMINARY; PRT; 723 AA.
ID Q8NBB9
AC Q8NBB9
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ3731.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wgatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK091050; BAC03574.1; -.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
KW Hypothetical protein.
SQ SEQUENCE 723 AA; 78133 MW; B279155820EA99AA CRC64;

Query Match 8.4%; Score 240; DB 4; Length 723;
Best Local Similarity 22.6%; Pred. No. 7.1e-09;
Matches 132; Conservative 54; Mismatches 198; Indels 200; Gaps 27;

Qy 19 SAEPLSAPISMPMPQYTMQPVYSQPHLPLQPHHSQSPHSPHSPHSPHSPHSPH 78
Db 84 SMPEPFVYSQMRPPLTMEP--PVLPSLPPVPMP--PALPATVPPPGMPP----- 131
Qy 79 PASSHDVYASSAAPIMPTTGVSLPPTSLFSLHPNPQAQAQAQSPHYPHPPH-----SV 131
Db 132 -----PVMPPLSLTSPVPPGM-----PPLSSAGPPFVLPPPSLFSAGPPPV 173
Qy 132 LPPASSACSYPQIAPAPRRDRADFNGLPSGAFSYSDGKPGQWDVVAANGAAPYCGKD 191
Db 174 LPPSLSLSTAPPVMPPLPPLS-SATPPPGIP-----PPGVQILPPOLT--NADVPFAS 224
Qy 192 SPRTQVVGSGRRGILPSVFRATPTVNGVNGTKNTTI-----PAKDADGKFPCHN 243
Db 225 SSQSSQVPEKPRPALP-----TPVS---FGSAPPTTYHPPLOSAGPSEQVNSKAP--- 272
Qy 244 CNKTYLHAKHLKRLHRTGDRPVMVLCVDTFSRDLKRLHFKQKCSIRGNTGTHLS 303
Db 273 -----LSKSL-----PY-----SSFSDDQL----- 289
Qy 304 HPNAHVRSQQAANPVKQVQD---EVSSVTPVPNGIPGTYGEG----- 346
Db 290 -----GESSAASQPIITAVKMPVRSGLLPPD---PRSSYLESPRGPRDGPREF 339
Qy 347 -----AVNGNGLAPAPPGYADHQTWG-----FPMSSVNG 375
Db 340 LGSRCGRPKGRPEGRNPDGPRVEGHPAECTKSKWGMIPRGASQFVITDSTSLSP 399
Qy 376 MGRGQ-----PEDAPPGRPHQAPWQAPKQSPYLVQPGADPSGHQ-----LNID 421
Db 400 RQSGPQWKGPAP--GQHQQQQPKQAEPISGN-KPFLADTSSNQKNKFMQSAAFSIA 456
Qy 422 RNIEQVK-----QPVQDPEKPVMPGHPGHELDW-TSMFQPAQEGYMFSSQSPG 472
Db 457 ADKDVKAASQENLSDSQEPKSEVSEGVSEGVSNWQNVQSMETQIDKQAVTQPVPL 516
Qy 473 GQEPHIAHVETERKYPTTTAGQSGMNGLYASTMSGDGTVPQ 516
Db 517 ANKVPFA-----QSTFESKTCGMEGGT--AVATSSLTADNDFKP 553

RESULT 5

Q8NPF45 PRELIMINARY; PRT; 1766 AA.
ID Q8NPF45
AC Q8NPF45
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE FLJ00353 protein (Fragment).
GN FLJ00353
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]


```

RESULT 7
Q9VQ94 PRELIMINARY; PRT; 1193 AA.
ID Q9VQ94 Q96016;
AC Q9VQ94; Q96016;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE CGL10882 protein (LP05220P).
GN CGL10882 OR BCJNA:LP05220.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatiades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beran B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananatiades P.G., Brandon R.C., Rogers Y.,
RA Banzon J.W., Center A., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."

```

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnick S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003583; AAF51283.2;
DR EMBL; AY052042; AAK93466.1;
DR FlyBase; FBgn0031408; CG10882.
DR GO; GO:0003824; Fcatalytic activity; IEA.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR007123; Gelsolin.
DR InterPro; IPR006900; Sec23 helical.
DR InterPro; IPR006896; Sec23 trunk.
DR InterPro; IPR006031; XYPPX.
DR InterPro; IPR006895; zf-Sec23_Sec24.
DR Pfam; PF00626; Gelsolin; 1.
DR Pfam; PF04815; Sec23 helical; 1.
DR Pfam; PF04811; Sec23 trunk; 1.
DR Pfam; PF02162; XYPPX; 14.
DR Pfam; PF04810; zf-Sec23_Sec24; 1.
DR PROSITE; PS00879; ODR_DC_2.2; 1.
DR PROSITE; 1193 AA; 129270 MW; 7F799A99D3CA1B1A CRC64;
SQ
Query Match 8.3%; Score 237; DB 5; Length 1193;
Best Local Similarity 23.5%; Pred. No. 2.1e-08;
Matches 137; Conservative 34; Mismatches 155; Indels 256; Gaps 32;
QY 10 PFSTSL---PQGSAPLPSSA-----PISSMPMPQYTWQPY---FVSQPH 48
DB 8 PPPTQQQPPQFGAPPNSGGWPPQQQLPQQPQQQLPQQQQQQPQYGAAPPPTSAAS 67
QY 49 LPPLQPHHSQSAPH-----SYMGQPPYPDLNR-----YPASHDVYASSAAP- 92
DB 68 QPYLNGNYQQQLATNSGGLSVGGVGANPLKPLLPQGAAPAAAPPTGFNQNSNAAPP 127
QY 93 -----IMPHTTVGS-----LPPTSLFSLHPNPAQAQA---QOS-----PHY 126
DB 128 PTNNNNAAPGAPPPTQAGSYVNGALPSS-----TFQSVASGINQMSLNSATLAGLPHWP 182
QY 127 PPHSVLP-----PASSAQSYQPIAPAPPRDRADFNNGLPASGAFSYDGKP--QG 175
DB 183 PPKAATPGAAPQPPTPAAGS*SQ-ELPQGP-----LP-----GQPPFSG 223
QY 176 WDPVAANGAAPPYCKDSDPTQVVGSGRGILIPVPGRATPVYNGTGKNTTIPAKDA 235
DB 224 QIPTSQAPSPY-----GVFSSRPGGFQULPFGATPPTTTPG-----LPQQQ 266
QY 236 DGKFPNCPNCKTYLHAKHLKRLLRHTGDRPYMVCVKDTFSRSDILKRFQKCSIRGN 295

```

```
Db 267 QGIPP----- 271
QY 296 PTCATHLSHPNAHVKRSQQAANPVQDEVSVTPPPNGIPGTYTGEGAVNGNLAP 355
Db 272 -----LQPGFI-----PQQPGFPFQQGLPFLSQPLFPQ--PGAPYG-----AP 310
QY 356 ARPGYADHQTMGPFMSVNGMGRGQPEDAFPG-----GRPHQAGAPWPAQKQSP 404
Db 311 QQGGVSG-----GPP-----GQAPGFPAGPPLPQQAAPQFGAPQGYGQGP 357
QY 405 -XLVQFGADP-----SGHLNIDRNIQVQVVDKPRPYMP--GHGPHGELDWT 454
Db 358 GYPPQPGQPMGYPYPPQCGQLG-----GCGYPPQCGAGPFGQGRPGFN- 402
QY 455 MQQPQAP-EGYMFQS-----MPGQBEIHAHVETER 485
Db 403 --QPPMEGAGNYQQAQPARRLDPQMP-----NPIQVMENQR 439
```

RESULT 8

```
Q9C170
ID Q9C170 PRELIMINARY; PRT; 722 AA.
AC Q9C170;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Stel2-like transcription factor.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Bobrowicz P., Ebbole D.J.;
RT "Neurospora crassa Stel2-like gene.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY027529; AAK1481.1; -
DR HSP; P08047; ISP2.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003120; TF_STE.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF02200; STE; 1.
DR Pfam; PF00036; zf-C2H2; 2.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00424; STE; 1.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW Metal-binding; zinc; Zinc-finger.
SQ SEQUENCE 722 AA; 79712 MW; 64B98CFDFC70A51 CRC64;
```

```
Query Match 8.1%; Score 233; DB 3; Length 722;
Best Local Similarity 21.1%; Pred. No. 2.2e-08;
Matches 104; Conservative 48; Mismatches 154; Indels 188; Gaps 21;

QY 12 STSLPQGSALPLS-----APISMPMPQTMQPY-----PVSQPH- 47
Db 264 SQTMSQSMALSLDGLAMVSYGAMTGMFAMTCPPQPMVKREPDYGRVQYQNGVPIQVHQ 323
QY 48 -----TLPLPQPHSQAAPHSYMGQPPY--RPD 74
Db 324 RHTSMPAYGLECSAPSFVSSHFDYSGRGLSFELTPHQOM-----GVIGEPAYIANEE 379
QY 75 LNRYPASSHDVYASSAAPIMHTTGVSLPPTSL--SHNPQQAQAQQ----- 121
Db 380 TGLYTAIPDLNAGALNGM-----QLPPSNLAAPSYPRSYAGSSISQDHARGYSMAA 433
QY 122 -----SPHY-----PH-----SVLPASSAQSPQIAPAPPRDR 154
Db 434 YTVIEGSPTYKQRRRRSIPPSVLSTATTAAAGTGVYKPSDLRRSVASGYPVAEGDES 493
```

```
QY 155 AD-----FNN-----GLPSGAFSYSDGKP-----QGWDP 178
Db 494 ADNSPPGLTYSNFALNMARQOHHEMLELSRHCTP---LSTVSGSPALNPMNISMGSQSYGP 550
QY 179 VAANGAAPYPGKDSRPTQVVGQGRGILPSVPGRATPVTVNGVGTGKNTTIPAKDADGK 238
Db 551 LNDDDELAPMSERQM-----SQGHFQVI-----RRARSAT--VSELG-----PYPKSHS 592
QY 239 FFCPNCKNTYLHAKHLKRHLRHTGDRPYMVCVLCKDTSRSDILKHFQKCSIRRGNPTG 298
Db 593 CFIPIGCGRVFKLEHLKRHVTRHTQERPVVCSYCKAFSRSDNLAQ----- 638
QY 299 ATHLSHPNAHVKRSQQAANPVQDEVSVTPPPNGIPGTYTGEGA-VNNGNLAPAR 357
Db 639 -----HKRTHDRGDGSGTNSLSGEEERYSS-----PTSEGGVYHAGSGAPNGNSSTPQS 689
QY 358 PGYADHQTMGFMS 371
Db 690 TWYNNLQTLSPMT 703
```

RESULT 9

```
Q8UZ11
ID Q8UZ11 PRELIMINARY; PRT; 3084 AA.
AC Q8UZ11;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL36 protein.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RA Klupp B.G., Fuchs W., Granzow H., Nixdorf R., Mettenleiter T.C.;
RT "The pseudorabies virus UL36 tegument protein physically interacts with the UL37 protein.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ421133; CAD19511.1; -
DR InterPro; IPR006928; Herpes_teg_N.
DR InterPro; IPR005210; Herpes_UL36.
DR InterPro; IPR002965; P-rich_extensions.
DR Pfam; PF04843; Herpes_teg_N; 1.
DR Pfam; PF03586; Herpes_UL36; 1.
DR PRINTS; PR01217; PRICHTEXTENS.
DR SEQUENCE 3084 AA; 324401 MW; ECAD9E1E3DC22D1A CRC64;
```

```
Query Match 8.1%; Score 232.5; DB 12; Length 3084;
Best Local Similarity 24.1%; Pred. No. 1.4e-07;
Matches 145; Conservative 41; Mismatches 245; Indels 171; Gaps 29;

QY 6 HPSRPSTSLPQGSALPLSAPISMPMPQTMQPYVPSQPHLPLPQPHHSQAPHSY 65
Db 2269 YPATAPAPETAPPEAPQAQPAAKKTP---QPQGP-----PPQPPSAQAQPAQ-- 2316
QY 66 MGQPPYRPLNRYPASSHDVYASSAAPIMHTTGVSLPPT---SFLSHPNPQAQAQAQS 122
Db 2317 --KPPAQP-----ATAAATTAPKATPQTPTPTAQTATAPPPPSATAAAQV 2361
QY 123 PHYP-----PPHSVLPPASSAQ--SYQPQIA-PAPPRDRADFNGLPSGAFSY 168
Db 2362 PPQPPSSQAAPKPRGAPPAPPAPPPSAQTTLPRPAAPPAPP-----PSAQTTL 2411
QY 169 SDGKPGQWDVVAANGAAPPKGDSPRTQVVGSGRRGILPSVPGRATPVTVNGVGTGK-- 226
Db 2412 PRPAP---PPPSAATAPTTPPAFGPAPSAKSKSDGRIVEPK--AGAPPDVRDAKFGKVA 2466
QY 227 NTIIPAKDAD-----GKFPCCPNKNTYLHAKHLKRHLRHTGDRPYMVCVLCKDTFSRSD 280
Db 2467 GAASPGRLDDERYKAGR--ADKCEHTVGDQHRTRSPPGTSDSEP-----ASASSSAAS 2519
```

```

QY 281 ILKRFQKCSI-----REGNPTGATHLS-----HPNAHVRSQ 313
Db 2520 TPREAASSSSPATPSDARGGIPSPSAGAAARAPGAPSGGRRGGGGAADASRGD 2579
QY 314 QAAANPVKPVQDEVSSVTPPPNGIPGTYG---EGAVNGNGLAPARPGYADHQTMGFPM 370
Db 2580 QPAGHLPAAPVAQPLPSAVLRKA--RTSVGIRDHSRRKSMALVPRRPVSPAPRQTGGLPP 2637
QY 371 SSVNGMGRGQPDAPFGG---RPHQ-----GAPWQAPKQSPYLVQPGADPSGH 416
Db 2638 AA-----QPGRPEGGLEPRPPOQAQAPATAPTGPAAAAATAAAPPVAPGPGG- 2689
QY 417 QLNIDRNIEQVKQV---VQDKRPVMPGH-----PGHEGLDWTSMFQQAPEGYMSQ 468
Db 2690 -----PVRRAGRPFA--GHLRHIIPPPQVQLEGVVVLPASPETPAPAQ 2734
QY 469 SMP-----GGPPIHAHVETER-----KYPTTTAGQESGMNGLYLASTMSGDGTQV 515
Db 2735 TQPPRSLAAPTSLAAPSEIERPAASAAAAATTTTSSSS-----ATTSSAPAAP 2787
QY 516 PA 517
Db 2788 PA 2789

RESULT 10
Q924S6
ID Q924S6 PRELIMINARY; PRT; 726 AA.
AC Q924S6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Zinc finger protein 219.
GN ZFP219 OR 2010302A17RIK OR ZNF219.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper selected cDNAs to
RT prepare full-length cDNA libraries for high-rate new gene discovery.";
RL Genome Res. 10:1617-1630 (2000).
DR EMBL; AB063578; BAB61057.1; -.
DR MGD; MGI:1917140; Zfp219.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; ZF-C2H2; 8.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
KW Metal-binding; Zinc; Zinc_finger.
SQ SEQUENCE 726 AA; 77857 MW; 82A51328FB6B111E CRC64;

Query Match
Best Local Similarity 23.6%; Pred. No. 2,9e-08;
Matches 128; Conservative 48; Mismatches 202; Indels 165; Gaps 28;

QY 16 PGGSAPLPSAPISMPMPQYTMQPVPSQPHLPPLQPHHSQSPAPHSYMGQPPYRDL 75
Db 222 PLAASTPEPPPPQPEPRSALEPE-PEPEPRPEPDREANPAPTAPPE---EPATPEF 277
QY 76 N-----RYFASHSDVYASSNAPI-----MHPTVGSILPPTS 106

```

```

Db 278 RCQVCGSQFTQSWFLKGHRKHKASFDHACVCGRCFKPEWFLKNHMKVHTS--KLGP-- 333
QY 107 FLHPNP-QAQAQOQSPH-----YPP--PHSVLPPASSAQSYQOP-----IAPA 148
Db 334 -LRAPGGSAPARAPQPPDLSLAYEPLGALLAPAPAPARERPPSLLGYLSVRAGEV 392
QY 149 PPRRRADFNGLPSGAFSYSDGKP-----QGWDFVA 180
Db 393 RPNCEGAD-----PGGSRSYGFRPLPSALPNRARHRHTEPEEBEEVVEEESWARGR 447
QY 181 ANGA--PYPGKDSR--TQVGSQRRGILPSVPGRATPVNG--VNGTKNTTIPAK 233
Db 448 SLGSLTSLHPNPGGSGQAPAAQTQAR-----STATQENGLLVGCTSEA---GR 496
QY 234 DADGKPCPCNCKTYLHAKHLRHLRHTGTRPYMCVLCCKDTFSRDLKRHFQKCSIR 293
Db 497 GATGK-DCFFCGKGRSAHLKVLRLVHTGERPKPCPDYAGTQSGSLKYHLQRHHRQ 555
QY 294 GNPTGATHLSPNAHVRS--QQQAANPVKPVQDEVSTVP--PPNGIPGTYGEGAVN 349
Db 556 RSSAGPGEPPPPPPPSQSGSLQPSGAKPTQASATWVEGTASTREPPSSSTGP----- 607
QY 350 GNGLAPARPGYADHQTMGFPMSVNGMGRGQPED---AFPGGRP-----HO----- 392
Db 608 GSRRKPASPGRTLNRG-----GEAPLDSLRAFGGEGAGAGALHRCULFCPPA 658
QY 393 -GAPWQA-----PKQSPYLVQPGADPSGHLNIDRNIEQVKQPVVQDKRPVY--PG 442
Db 659 TGAPELMALHQLVHHSRRARGRQPRADTSP-----TVRAPSGETPPSPLEEBE 709
QY 443 HPG 445
Db 710 SPG 712

RESULT 11
Q9CVF3
ID Q9CVF3 PRELIMINARY; PRT; 596 AA.
AC Q9CVF3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 2010302A17RIK protein (Fragment).
GN ZFP219 OR 2010302A17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Yuzhakov-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:695-690 (2001).

```

```

DR EMBL; AK008516; BAB25714.1; -
DR MGD; MGI:1917140; Zfp219.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; PF00096; zf-C2H2; 6.
DR Pfam; SM00355; Znf_C2H2; 7.
DR PROSITE; P500028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; P50157; ZINC_FINGER_C2H2_2; 4.
KW Metal-binding; Zinc; Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 596 AA; 63571 MW; 49C750FDC6760DFE CRC64;

Query Match
Best Local Similarity 8.0%; Score 230.5; DB 11; Length 596;
Matches 128; Conservative 48; Mismatches 202; Indels 165; Gaps 28;

QY 16 PQSAPLPSAPITSSMPMPQVTPQVPSQPHLPQLPHHSQSPAPHSYMGQPYRPLD 75
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92 PLAASTPPEPPPPQCEPSALEPE-PEPEPRPEPDREANPAPTAPPE---EPPAPPEF 147
QY 76 N-----RYPASSHDVYASSAPI-----MPHTTVGSLPPTS 106
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
148 RQVCQSQSTQSWFLKGHRMKKASFDHACPVCGRCFKEPFWFLKNMKVHTS--KLGP-- 203
QY 107 FLSHNP-QAQAQAQSPH-----YPP--PHSVLPASSAQSYQPQ-----IAPA 148
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 -LRAPGGSAPARAPQPPDLSLAYEPLGLPALLAPAPAEERPPSLLGYLSVRAGEV 262
QY 149 PPRDRADFNNGLPSPGAFYSQDKP-----PGRGRSYGSPRPLPSALPNRARRHRTPEEPEEVEVVEABEESWARGR 317
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 ANGAA---PYPGKDSPR-TQVVGSGRGRLPVSFGRTPTNG--VNGTGKNTIPAK 233
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318 SLGSLTSLHPNPGEGSGQAPAAAGTQAR-----STATQENGLLVGTTRSEA---GR 366
QY 234 DADGKFPCCNCKNTYLHAKHLKHLRHRTGDRPYMCVLCKDTFSRDLKRRHFQKGSIR 293
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 GATGK-DCPFCGKSFSAHHLKVELRVHTGERDYKCPHCDYAGTQSGSLKYLQRRHRE 425
QY 294 GNPTGATHLSHPNAHVKS--QQQAAANPKPVQDEVSSIVP--PENGIPGTYGGAVN 349
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 RSAGPGPPPPPPPPSQRGSLQPSGAKPTQASATWVGFTASTRPPSSSTGP----- 477
QY 350 GNGLAPARPGYADHQTGMFPMSSVNGMGQPED----APPGGRP-----HO----- 392
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 GSRKPAASPGRTLRNG-----GEAEPLDLSLRAGPGGEAGAGNLRCLCFPA 528
QY 393 -GAPWPQA-----PKQSYLVQPGADPSGHQNLNDRNIEQYQPVQDPKRPVM--PG 442
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
529 TGAPELMALHLQVHSGRRARRGRQPRADTSP-----TYVRAPSGETPPSPLEEEG 579
QY 443 HPG 445
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
580 SPG 582

RESULT 12
Q9W3G1 Q9W3G1 PRELIMINARY; PRT; 926 AA.
AC Q9W3G1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CGL0555 protein.
GN CGL0555.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

```

RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.,
RA Arlil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strongski M., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu Q., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003443; AAF46366.1; -
DR Flybase; FBgn0030034; CGL0555.
DR InterPro; IPR007726; SSXT.
DR Pfam; PF05030; SSXT; 1.
SQ SEQUENCE 926 AA; 93004 MW; 6FD5B9F77C36C006 CRC64;

Query Match
Best Local Similarity 8.0%; Score 229.5; DB 5; Length 926;
Matches 142; Conservative 44; Mismatches 229; Indels 205; Gaps 30;

QY 4 RNHP-----SRPPTSLP-QGSAPLPS-----APISMPMPQVMTQP 39
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
351 QCHPVYRNAQGGQPGAGQVPGQGGQGVQSVINPNAAPQQRPNNGFLSQPNPQQQQQ 410
QY 40 QYFVSQPHLPLOPHHSQSPAPHSYMGQPYRPLNRYPPASSHDVYASAPIMHTTV 99
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
411 PQFGGQQ---PENNQQQQQQQTGP---GGFGQP-----GAGGEGVP----- 445
QY 100 GSLPPTSLSHENPQAQAQAQSPHYPP-----PHSVLPASSAQSYQPQIAPAPRD 152
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
446 ----PQSPYRVSVOQQQQQSHYFPFPQFQYQYQYQYQYQYQYQYQYQYQYQYQY 501
QY 153 RRADFNNGLPAGSFSYDQKPGWDFVAANGAAPYPKDSPTQVVGSGRGRLPSVPG 212
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
502 AQGGYHHG-PAGAATGASG--HGYQPNAGAGQGPFGAYPPPP---GSQQ-----VPPVPG 551
QY 213 RATPVTN-----GVNGTGKNTTIPAKDADGKF-PCPNCNKTLYLHAKHLKHLRHRTGDRPY 267
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
552 QQQPPPPPPPPGQPTGCGQQQPPGPPQSQYGYGPPPPQNS-----AGGPPPP 596
QY 268 MCVLCKDTFSRDLKRRHFQKGSIRRNPTG-----ATHLSHPNAHVKSQQQAAA 318
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 M-----GYAGYPPNPGYQQAAGAGGPPPPSGWPPPPPTSSAQSPYQAYQQQQQAAA 650

```



```
Db 337 MCGMGLPGPKGPTLAPSPHLPASS-----AFGPP--MRYPTSSSSSSAAAS 387
QY 168 YSDGKPGQMDPVAANGAAPYPGKDSRPTQV-VGSQGRRGILPSVPGRAT-----PVTN 219
Db 388 SSSSSSSASQYPASQALPSYPHSPFPPTSMVSQNPPKYTPQLPSQAVMSQGPDPDPY 447
QY 220 GUNGTGKNTTIPAKDADGKFPFCNCKNTYLHAKHLKRHLRLHRTGDRPYMCVLCCKDTFSRS 279
Db 448 G-RLGNNTHP-----GFPF-PTGQSTAHPAAPTHH---HQQQPQ-----485
QY 280 DILKRHFQKSIIRGN---PTCA-----TLSHPNNAHVKRSQQQAAANPVKPVQD 326
Db 486 ---QQH-----HGNHGAPPPGAYPHPLESSNSHAHP---YNNMPSLSGSLRPPYPP---530
QY 327 EVSSTVPPNGIPGTTYGEGAVNGGLAPRPGYADHQTMGFPMSSVNGMGRGQPEDAFP 386
Db 531 -GAHLPPPHG--QVSYNQAGNGP-----PVSSNSSG-SSSQASYS 569
QY 387 GRPH-----QGAWPQAFKQSPYLVPQADPSGHQINIDRNIETQVKPVVQDPKRPVMP 441
Db 570 CSHPSQSQPGQASYPFPF-----VPPVTTSSA-----TLSTVIATVASSPAGYKTA 616
QY 442 GHPGHPGELDTWTFQPEGYMSQSMGGOEPIHAHVETKRYPTTAGQESGMNG 501
Db 617 SPGPP-----QYKRAPSGSYKATPPGYKGGP-----PSFRTGTPPGYRG 660
QY 502 LYLATMSGDGVVQP 516
Db 661 ---TSPPAGGTFKP 672

RESULT 15
Q7Z718
ID Q7Z718 PRELIMINARY; PRT; 413 AA.
AC Q7Z718;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Specificity protein 7 short isoform.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Milona M.-A., Gough J.E., Edgar A.J.;
RT "Expression of human specificity protein 7 in osteoblasts.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY150674; AAN85557.1; --
SQ SEQUENCE 413 AA; 43092 MW; FD6F791FFDCB7073 CRC64;

Query Match 7.7%; Score 221.5; DB 4; Length 413;
Best Local Similarity 24.1%; Pred. No. 7.6e-08;
Matches 111; Conservative 30; Mismatches 139; Indels 181; Gaps 22;

QY 11 PSTSL-----PGSGAPLPSA-----PIS-SMPMPQYTMQCYPVSQPHLPLPQPHHS 57
Db 49 PFTSTNGLLSPAGSPAPTSYANDYPPFHSFGPFGPTGTQDFGLLVPKGHS-----S 100
QY 58 QSPAPHSYMGQPPYRFDLNRYPASSHDVYASSAAPIMHTTVGSLPPTSFSLSHP-----111
Db 101 SDCLFSVYTLDMTHPYGSWYKAGIH-----AGISGPP--GNTFTFWMDMHPGNGWLG 151
QY 112 NFQQAQAQOQ-----SPHYPPHSHVLPASSAQSYQPIAPAPRRDRADFNGLPSGAF 166
Db 152 GQGGQGGGLQGTLP TGPAQPLNPLP-----TYPSPDFAPLNP-----189
QY 167 SYSDGKPGQWDVPAANGAAPYPG-----KSPRTQVVGSGQR---RGILPSV 210
Db 190 -----APYPAPHLQPQGHVLPQDVYKPKAVGNSQLEGGSGGAKPP 231
QY 211 PGRATPVTVNGVTGKNTTIPAKDADGKFC-----P 242
```

```
Db 232 RGASTGGSGYGGSGA-----GRSSCDPCNQELERLGAAGLKKKPIHSCHIP 281
QY 243 NCKNTYLHAKHLKRHLRLHRTGDRPYMC--VLCKDTFSRSDILKRHQ-----KCSIR 292
Db 282 GCGKVGKASHLKAHLRWHHTGERPFVCNWLFCGKFRTRSDLELHRVHTHTREKKTCLC 341
QY 293 RGNPTGATHLS-HPNAHVKRSQQQAAANPVKPVQDVSSTVPPPPNGIPGTTYGEGAVNGN 351
Db 342 SKRFTSDHLSKHQTHGE-----PGGPPPSG--PKELGEGRSTGE 381
QY 352 GLAPARPGYADHQTMGFPMSSVNGMGRGQPEDAFPGRPHQ 392
Db 382 EBAQTTPR-----PSAS-----PATPEKA-PGGSPEQ 407

Search completed: April 1, 2004, 17:48:20
Job time : 49 secs
```